Crystal structure of Thermoplasma acidophilum citrate synthase

Russell, Rupert J. M.

Award date:
1994

Awarding institution:
University of Bath

CRYSTAL STRUCTURE OF *THERMOPLASMA ACIDOPHILUM* CITRATE SYNTHASE

submitted by Rupert J. M. Russell

for the degree of PhD

of the University of Bath 1994

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGEMENTS

I must initially thank my supervisors Dr. Michael Danson and Dr. David Hough and my honorary supervisor Dr. Garry Taylor for their valued help and encouragement throughout my PhD. I hope that my work in the past and future (yes, they gave me a job too) is reward enough for putting up with my sometimes 'too horizontal' a manner.

My thanks also go to my mum and dad for their love and support, Wendy and Richard for being such good friends and finally to Andrew and William for giving me the chance to regress and play with their toys.

There have been many people throughout my time here in Bath who have helped me both intellectually and socially. I thank you all.

My final thanks must go to SERC and Zeneca BioProducts (formerly ICI) for their financial support providing me with my much needed Grants©.

ABBREVIATIONS

Amino acids: Single and Triple Letter Codes

Alanine	ALA	A	Leucine	LEU	L
Arginine	ARG	R	Lysine	LYS	K
Asparigine	ASN	N	Methionine	MET	M
Aspartic acid	ASP	D	Phenylalanine	PHE	F
Cysteine	CYS	C	Proline	PRO	P
Glutamine	GLN	Q	Serine	SER	S
Glutamic acid	GLU	E	Threonine	THR	T
Glycine	GLY	G	Tryptophan	TRP	W
Histidine	HIS	H	Tyrosine	TYR	Y
Isoleucine	ILE	I	Valine	VAL	V

Other Abbreviations

AMP, adenosine monophosphate

ATP, adenosine triphosphate

CD, circular dichroism

Citrate synthase or CS, citrate synthase (EC 4.1.3.7)

CoA, coenzyme A

DTNB, 5, 5' -dithio-bis-(2-nitrobenzoic acid)

EDTA, (disodium) ethylenediaminetetraacetate

FPLC, fast protein liquid chromatography

Gdn, Guanidine

IEF, isoelectric focussing

M_r, relative molecular weight

NADH, nicotinamide-adenine dinucleotide

NADPH, nicotinamide-adenine dinucleotide phosphate

NCS, Non-crystallographic symmetry

OAA, oxaloacetate

PAGE, polyacrylamide-gel electrophoresis

PBE118, Polybuffer exchanger 118

PEG, polyethylene glycol

RMS, Root mean square

rRNA, ribosomal RNA

SDS, sodium dodecyl sulphate

Tris, Tris(hydoxymethyl)methylamine

w/v, Ratio of weight to volume (g/ml)

SYMBOLS

E_A, Energy of activation

k, rate constant

R, Gas constant

a, b, c, α, β, γ, Real space cell axis lengths and interaxial angles

B, Isotropic temperature factor

$\mathbf{F}_{obs}$, F_{obs}, Observed structure factor and its amplitude

$\mathbf{F}_{calc}$, F_{calc}, Calculated structure factor and its amplitude

f_j, Atomic scattering factor

h, k, l or $\mathbf{h}$, Miller indices

R, Crystallographic residual

V, Unit cell volume

w, weighting parameter

x, **y**, **z**, Cartesian coordinate axes

α, phase angle

α, β, γ, Eulerian angles

ϕ, ψ, κ, Spherical polar angles

λ, wavelength

σ, standard deviation

fom, figure of merit

ABSTRACT

In order to gain detailed structural insights into features conferring thermal stability of Archaeal enzymes, the gene for citrate synthase from the thermoacidophilic Archaeon *Thermoplasma acidophilum* had previously been cloned, sequenced and over-expressed in *Escherichia coli* (Sutherland *et al.*, 1990, 1991). The recombinant protein has been purified to homogeneity by a heat denaturation step at $65°C$ followed by one of two different chromatographic techniques - chromatofocussing or affinity chromatography using Matrex Red GelA.

The purified protein was used for wide spread crystallization trials and three different crystal types were obtained; two belonging to the orthorhombic crystal system (types 1 and 3) and one to the monoclinic crystal system (type 2). Data sets were collected for each crystal type on the in-house area detector, with crystal type 2 diffracting to a resolution limit of 2.4Å.

Molecular replacement techniques were employed on each data set with pig heart citrate synthase used as a search model. After the use of many different molecular replacement packages and various deletion search models, a solution for crystal type 2, which contained two dimers in the asymmetric unit, was identified using AMORE. A screen of heavy atoms was also undertaken in the search for isomorphous derivatives, but no useful phase information was gained from this study.

The structure for *Tp.acidophilum* citrate synthase has been partially refined using a simulated annealing protocol within X-PLOR, and has an R-factor of 23.5%. Structural comparisons have been made between this thermostable citrate synthase and the mesophilic pig heart citrate synthase. A number of observations have been made that may be responsible for the thermostability of *Tp.acidophilum* citrate synthase, which has been

characterized using thermal inactivation assays and circular dichroism spectroscopy.

CONTENTS

		PAGE
TITLE		
ACKNOWLEDGEMENTS		
ABBREVIATIONS		
ABSTRACT		
Chapter 1	**INTRODUCTION**	1
1.1	CITRATE SYNTHASE	1
1.1.1	The citric acid cycle	1
1.1.2	Biological role of citrate synthase	1
1.1.3	Diversity of citrate synthases	3
1.1.4	Regulation of citrate synthase activity	3
1.1.5	3-dimensional structure of citrate synthase	5
1.1.6	Proposed mechanism of action of citrate synthase	8
1.2	CLASSIFICATION OF LIVING ORGANISMS	11
1.2.1	Woese and his three domains of life	11
1.2.2	Lake and his five groups of life	11
1.3	PHYLOGENETIC ANALYSIS OF LIVING ORGANISMS	11
1.3.1	The founding father's view - Woese and his rooted three domains	11
1.3.2	The old pretender's view - Zillig and his fusion event	14
1.3.3	The black sheep's view - Lake and his eocyte tree	14
1.3.4	Why the confusion?	18

1.3.5	Why choose the Woese convention?	18
1.4	THE ARCHAEA	19
1.4.1	Biotechnological potential of the Archaea.	19
1.5	*THERMOPLASMA ACIDOPHILUM*	20
1.5.1	Classification of *Tp.acidophilum*	20
1.5.2	Environmental adaptation of *Tp.acidophilum*	20
1.5.3	*Tp.acidophilum* citrate synthase	23
1.6	PROTEIN STABILITY	24
1.6.1	Introduction	24
1.6.2	Primary sequence comparisons	25
1.6.3	Structural analysis of homologous mesophilic and thermophilic proteins	26
1.6.4	Mutagenesis experiments	28
1.6.4.1	Structural context	28
1.6.4.2	Hydrogen bonds	29
1.6.4.3	Hydrophobic interactions	29
1.6.4.4	Conformational entropy	30
1.6.4.5	Disulphide bonds	31
1.6.4.6	α-Helical stability	31
1.6.4.7	Specific examples	32
1.6.5	Concluding remarks	33
1.7	BACKGROUND AND AIMS	34
Chapter 2:	**Experimental Techniques and Programs**	35
2.1	MATERIALS	35

2.1.1	Enzymes, reagents and other materials	35
2.2	METHODS	35
2.2.1	Assay for citrate synthase activity	35
2.2.2	Bradford protein concentration determination	35
2.2.3	Production of *Tp.acidophilum* citrate synthase in *E.coli*	36
2.2.4	Preparation of cell extracts from *E.coli*	36
2.2.5	Heat purification	36
2.2.6	Gel filtration	36
2.2.7	Chromatofocussing	37
2.2.8	Matrex Red GelA affinity chromatography	37
2.2.9	Crystallization trials	37
2.2.10	Thermostability assay	37
2.2.11	Circular dichroism spectra	38
2.3	PROGRAMS USED	38
Chapter 3:	**Purification of *Tp.acidophilum* citrate synthase**	41
3.1	INTRODUCTION	41
3.2	RESULTS	41
3.2.1	Thermal denaturation of host *E.coli* proteins	41
3.2.2	Purification of *Tp.acidophilum* citrate synthase using chromatofocussing	41
3.2.3	Purification of *Tp.acidophilum* citrate synthase using Dye-ligand chromatography	42
3.3	DISCUSSION	47
3.3.1	Chromatofocussing	47

3.3.2	Dye ligand chromatography	48
3.3.3	The better of the purification procedures?	49
Chapter 4:	**Characterisation of *Tp.acidophilum* citrate synthase**	50
4.1	INTRODUCTION	50
4.1.1	The Arrhenius plot of thermal inactivation	50
4.1.2	Circular dichroism spectroscopy	51
4.2	RESULTS	52
4.2.1	Thermal inactivation of *Tp.acidophilum* citrate synthase	52
4.2.2	Circular dichroism spectroscopy studies *Tp.acidophilum* citrate synthase	60
4.2.3	Thermal denaturation studies of *Tp.acidophilum* citrate synthase	62
4.2.4	Guanidine-HCl denaturation of *Tp.acidophilum* citrate synthase	65
4.2.5	Guanidine-HCl inactivation of *Tp.acidophilum* citrate synthase	70
4.3	DISCUSSION	70
Chapter 5:	**Crystallization, Data Collection and Analysis**	75
5.1	X-RAY DIFFRACTION THEORY	75
5.1.1	Diffraction from a crystal	75
5.1.2	Bragg's Law	75
5.1.3	The Laue equations	75
5.1.4	Calculation of Structure factors and Electron density	76
5.2	CRYSTALLIZATION INTRODUCTION	78
5.3	RESULTS	79

5.3.1	Crystallization trials	79
5.4	CRYSTALLIZATION DISCUSSION	81
5.4.1	Differences between crystals grown from Red GelA and chromatofocussing purifications	84
5.5	DATA COLLECTION AND ANALYSIS	85
5.5.1	The Area Detector	85
5.5.2	Data Collection	86
5.5.3	Data Processing	89
5.5.4	Space Group Determination	92
5.5.5	Data Processing Statistics	94
5.6	DATA ANALYSIS OF SUBSTRATE BOUND CRYSTALS	97
5.7	DATA COLLECTION/ANALYSIS DISCUSSION	97
Chapter 6:	**Molecular Replacement**	100
6.1	MOLECULAR REPLACEMENT THEORY	100
6.1.1	Rotation functions	100
6.1.2	Translation functions	102
6.2	INTRODUCTION	105
6.3	RESULTS	106
6.3.1	Data sets used	106
6.3.2	Search models used	106
6.4	THE SELF ROTATION FUNCTION	108
6.5	THE CROSS ROTATION AND TRANSLATION FUNCTION	110

6.5.1	MERLOT	110
6.5.2	X-PLOR	111
6.5.3	AMORE	114
6.6	DISCUSSION	122
Chapter 7:	**The Search for Heavy Atom Derivatives**	126
7.1	INTRODUCTION	126
7.2	RESULTS	127
7.3	DISCUSSION	131
Chapter 8:	**Refinement of *Tp.acidophilum* citrate synthase**	133
8.1	REFINEMENT THEORY	133
8.2	THE X-PLOR ENERGY FUNCTION	134
8.2.1	The Empirical Energy Function	134
8.2.2	The Effective Energy Function	135
8.3	MOLECULAR DYNAMICS	136
8.4	RIGID BODY REFINEMENT	138
8.5	REFINEMENT STRATEGY	138
8.5.1	Rigid Body Refinement	138
8.5.2	Refinement by Simulated Annealing	138
8.6	REFINEMENT PROGRESSION	141
Chapter 9:	**Structural Characterisation of *Tp.acidophilum* Citrate synthase**	157
9.1	INTRODUCTION	157
9.2	OVERALL FOLDED CONFORMATION	157
9.2.1	Active site conformation	161

9.3	POSSIBLE FEATURES CONFERRING THERMOSTABILITY	166
9.3.1	N-terminal deletion	166
9.3.2	Loop deletions	166
9.3.3	Subunit-subunit interactions	170
9.3.4	Helix-capping residues	178
9.3.5	Active site conformation/environment	181
9.4	POSSIBLE ARCHAEAL PROTEIN FEATURES	182
9.5	CONCLUSION	183
Chapter 10:	**Conclusions and Further work**	184
REFERENCES		187

Chapter 1 : Introduction

Citrate synthase, a key metabolic enzyme from the thermoacidophilic Archaeon *Thermoplasma acidophilum*, has been targetted for use in comparative structure/function studies to gain insights into both the evolution of organisms and protein thermostability. Therefore, an introduction to the enzyme will be followed by a brief synopsis about the current understanding of the phylogenetic relationships between organisms, highlighting the existence of the Archaea. Finally a brief review of the proposed basis for protein thermostability will be presented.

1.1 CITRATE SYNTHASE

1.1.1 The citric acid cycle

The citric acid cycle has two essential functions in living cells: (i) to oxidize metabolites and produce NAD(P)H , leading to the generation of energy and, (ii) to produce metabolites necessary for biosynthesis. The main steps of the pathway are summarised in Fig. 1.1. Organisms inhabit a diverse range of environments and this leads to differing emphases being placed on the role played by the citric acid cycle in specific organisms. Citrate synthase is usually classed as the 'first' enzyme of the cycle and thus it plays an important role in the regulation of this metabolic pathway. The enzyme is discussed in more detail below.

1.1.2 Biological role of citrate synthase

Citrate synthase plays a vital role in central metabolism, facilitating the entry of acetyl-CoA into the citric acid cycle. It catalyses the reaction :

$$\text{oxaloacetate} + \text{acetyl-CoA} + H_2O \text{ <—> } \text{citrate} + \text{CoA} + H^+$$

thus forming citrate by the creation of a carbon–carbon bond. The overall

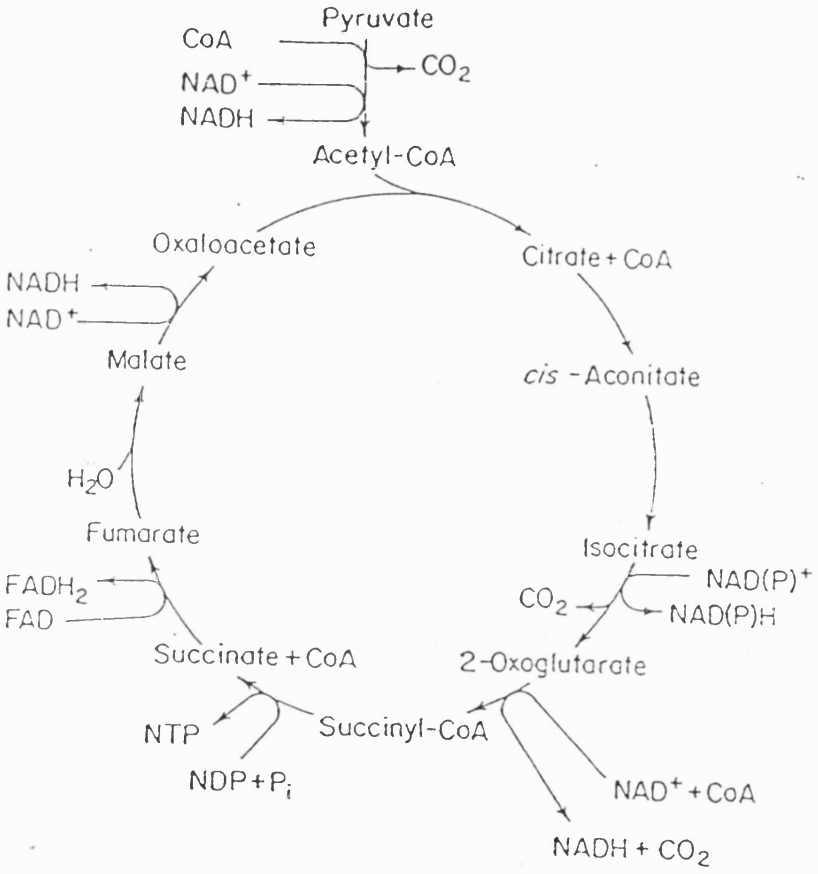


Figure 1.1: The Citric acid cycle

reaction can be divided into three sub-reactions, namely (a) enolisation: coordinated acid-base abstraction of a methyl proton from acetyl-CoA to produce a carbanion; (b) condensation: coordinated acid-base nucleophilic attack of the carbanion of acetyl-CoA on the carbonyl group of oxaloacetate to produce citryl-CoA; and (c) hydrolysis: hydrolysis by an activated water of citryl-CoA to produce citrate and CoA (Remington, 1992).

1.1.3 Diversity of citrate synthases

Citrate synthase is present in essentially all living organisms and has been studied from organisms representing the three domains of life. There are two basic forms of citrate synthase, a dimeric 'small' form found in Eucarya, Gram-positive bacteria and Archaea and a hexameric 'large' form found in Gram-negative bacteria. Both forms are made up of identical subunits of M_r values of approximately 50 000, the hexamer existing functionally as a trimer of dimers (Else *et al.*, 1988). 14 citrate synthase sequences are thus far known (see Table 1.1).

There is a high degree of sequence identity between citrate synthases within a domain, with eucaryotic and bacterial citrate synthases showing mean sequence identities of 53% and 62% respectively, but sequence identities between domains is far lower (18%–28%). *Thermoplasma acidophilum* citrate synthase has a mean identity of 28% with bacterial citrate synthases and 20% with eucaryotic citrate synthases.

1.1.4 Regulation of citrate synthase activity

Since citrate synthase is the 'first' enzyme of the citric acid cycle, its regulation is thought to be of the upmost importance in the control of metabolic flux through the cycle and so it has been studied widely. The hexameric form and the dimeric form have distinct regulatory properties.

Table 1.1: Primary amino acid protein sequences of citrate synthases.

ORGANISM	CLASSIFICATION	REFERENCE
Pig heart	Eucarya	Bloxham *et. al* (1981)
Saccharomyces cerevisia (mitochondria)	Eucarya	Suissa *et. al* (1984)
Saccharomyces cerevisia (glyoxysome)	Eucarya	Rosenkrantz *et. al* (1986)
Arabidopsis thaliana	Eucarya	Unger *et. al* (1989)
Tetrahymena thermophila	Eucarya	Numata *et. al* (1991)
Escherichia coli	Bacteria	Bhayana and Duckworth (1984), Ner *et. al* (1983)
Rickettsia prowazekii	Bacteria	Wood *et. al* (1987)
Acinetobacter anitratum	Bacteria	Donald and Duckworth (1987)
Pseudomonas aeruginosa	Bacteria	Donald *et. al* (1989)
Acetobacter aceti	Bacteria	Fukaya *et. al* (1990)
Coxiella burnettii	Bacteria	Heinzen *et. al* (1991)
Mycobacterium smegmatis	Bacteria	David *et. al* (1991)
Bacillus sp.strain C4	Bacteria	Schendel *et. al* (1992)
Thermoplasma acidophilum	Archaea	Sutherland *et. al* (1990)

The large hexameric citrate synthase of Gram-negative bacteria is allosterically inhibited by NADH (Weitzman,1981 and references therein), but the small dimeric form present in all other organisms is isosterically inhibited by ATP and not allosterically inhibited by NADH (Weitzman and Danson, 1976 and references therein). Inhibition by these nucleotides can be seen as a form of negative feedback control of energy production since NADH is the primary end product of the cycle and ATP is the ultimate end product.

Other regulatory properties are seen in specific citrate synthases. AMP has been shown to reactivate the NADH-inhibited enzyme from obligately aerobic Gram-negative bacteria, but not the citrate synthase from facultatively anaerobic Gram-negative bacteria (Weitzman and Jones, 1968). This is probably due to the facultative anaerobes having the ability to produce energy by fermentation, with the obligate aerobes being dependent on precise control of the citric acid cycle for their energy production.

1.1.5 3-dimensional structure of citrate synthase

The crystal structures of four different crystal forms of pig and chicken heart citrate synthases have been elucidated for a total of 13 apo– and substrate-bound forms - Table 1.2. Each monomer of the dimer consists of two domains, a large and a small one, with the substrate binding site situated in the cleft between the two domains. Citrate synthase is almost entirely α–helical, the large domain containing 15 helices and the small domain containing 5 - Fig. 1.2. The two monomers are widely interdigitated, with residues from one monomer involved the the active site of the other monomer. The catalytic cleft is made up of residues from both the large and small domains. Citrate synthase exists as an open form, which on binding of substrates is converted to a closed form by a 18° rotation of the small domain with respect to the large domain (Remington *et al.*,

Table 1.2: Known crystal structures of citrate synthase.

ORGANISM	SUBSTRATE	ENZYME FORM	RESOLUTION (Å)	REFERENCE
Pig heart	—	OPEN	2.9	Remington *et. al* (1982)
Pig heart	citrate	OPEN	2.7	Remington *et. al* (1982)
Chicken heart	—	OPEN	2.8	Liao *et. al* (1991)
Pig heart	citrate	CLOSED	2.2	Wiegand *et. al* (1986)
Pig heart	citrate and CoA	CLOSED	2.0	Wiegand *et. al* (1986)
Chicken heart	citrate and CoA	CLOSED	1.7	Remington *et. al* (1982)
Chicken heart	citrylthioether CoA	CLOSED	2.5	Remington (1992)
Pig heart	oxaloacetate and S-ActCoA	CLOSED	2.9	Wiegand *et. al* (1984)
Chicken heart	oxaloacetate and CMCoA	CLOSED	1.9	Karpusas *et. al* (1990)
Chicken heart	L-malate and CMCoA	CLOSED	1.7	Karpusas *et. al* (1991)
Chicken heart	D-malate and CMCoA	CLOSED	1.9	Karpusas *et. al* (1991)
Chicken heart	L-malate and acetyl-CoA	CLOSED	1.9	Karpusas *et. al* (1991)
Chicken heart	D-malate and acetyl-CoA	CLOSED	1.9	Karpusas *et. al* (1991)

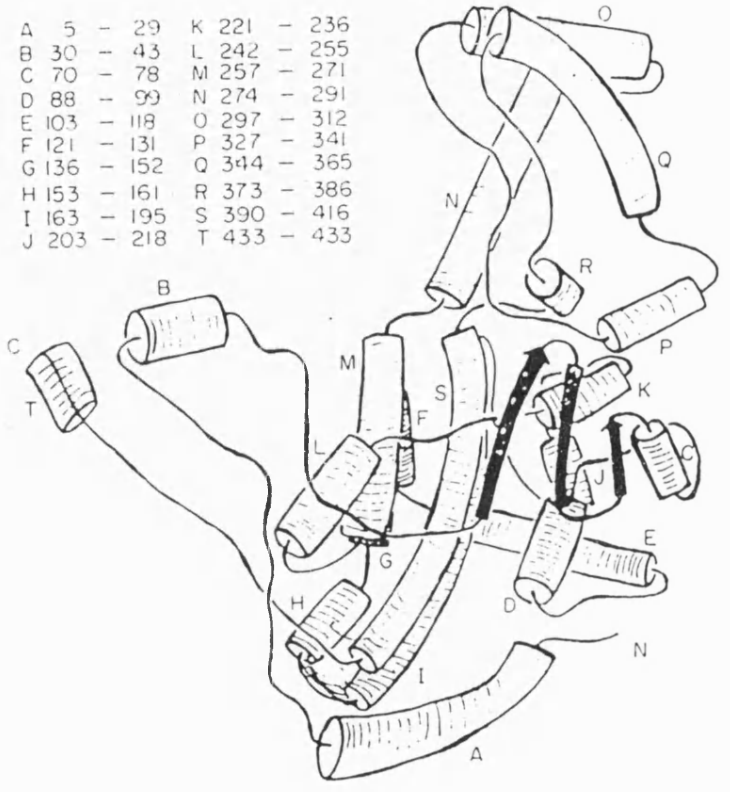


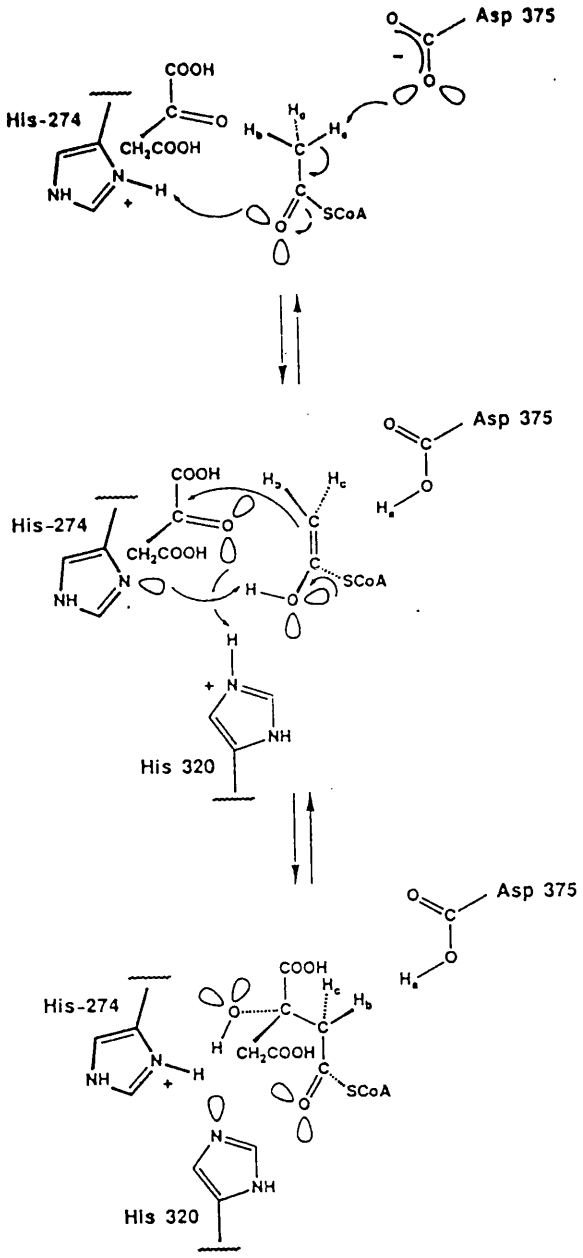
Figure 1.2: Schematic representation of a monomer of pig heart citrate synthase, looking down the two-fold axis. The 20 α-helices (A-T) are displayed as barrels. Reproduced from Remington *et al.* (1982).

1982). This large conformational change in structure has been proposed to be necessary to ensure precise placement of the key catalytic residues. The crystal structure of chicken heart citrate synthase has also been elucidated (Liao *et al.*, 1991) and shows an almost identical conformation to the pig heart enzyme. The RMS difference between the Cα atoms of the two enzymes is only 0.45Å. To date, no other 3-dimensional structures of citrate synthases have been elucidated.

1.1.6 Proposed mechanism of action of citrate synthase

Both detailed structural and experimental studies have led to the proposition that citrate synthase can exist in 'open' and 'closed' conformations. The open form facilitates substrate entry and product release and the closed form catalyzes sequentially all the steps of the reaction. The binding of substrates occurs in an ordered manner with OAA binding first, leading to a conformational change and the creation of the acetyl-CoA binding site. HIS274, ASP375 and HIS320 (pig heart CS numbering) have been shown crystallographically and experimentally to be the residues central to the catalytic action of citrate synthase (Alter *et al.*, 1990, Zhi *et al.*, 1991). These residues, and the ones forming the active site, are conserved in all known sequences, suggesting a universal mechanism of catalytic action. The key catalytic groups are on opposite sides of the catalytic cleft: HIS274 is on the large domain and HIS320 and ASP375 are on the small domain.

The rate limiting step of the reaction has been shown to be the formation of the enol intermediate of acetyl-CoA. Scheme 1.1 outlines the concerted acid-base mechanism of the enolisation reaction. HIS274 protonates the carbonyl oxygen of acetyl-CoA and due to the proximity of ASP375 to HIS274 which allows favourable electrostatic interactions and thus coupled



Scheme 1.1: Mechanism for the condensation reaction of citrate synthase. The first stage (enolisation) comprises the deprotonation of the methyl group of acetyl CoA by ASP375 with the concerted protonation of the carboxyl oxygen by HIS274, forming a neutral enol intermediate. The second stage (condensation) involves the rotation of the enol around the HIS274 H-bond and subsequent attack of the carbonyl carbon of oxaloacetate. HIS274 is then protonated in a concerted fashion with the protonation of the carbonyl oxygen by HIS320. Therefore both stages invoke coordinated acid-base catalysis. Reproduced from Remington (1992).

charged states, ASP375 deprotonates the methyl group at the same time to form the neutral enol intermediate.

The next step of the reaction mechanism is the condensation step. This again is thought to occur by a concerted acid-base catalysis. HIS320 acts as the acid and protonates the carbonyl oxygen of OAA, and HIS274 acts as the base and recycles the proton that it donated in the enolisation step. HIS274 is hydrogen bonded to the product complex and the enol rotates around this bond to attack the activated carbonyl of OAA to form the intermediate citryl-CoA.

The final step in the reaction mechanism is the hydrolysis of citryl-CoA to form citrate and CoA. This is probably the least well understood of the proposed reaction steps but two schemes have been proposed. Firstly (Scheme 1.2) Remington (1992) suggests that an as yet unidentified base deprotonates and thus activates a water molecule which attacks the thioester forming a tetrahedral intermediate. Again this is a concerted reaction involving the recycling of a proton to HIS274. Secondly (Scheme 1.3) Alter *et al.* (1990) propose that ASP375 forms an anhydride complex via attack of the nucleophilic carbon of citrate, thereby releasing CoA. The anhydride is then attacked by a base-activated water molecule to form citrate and restores the initial state of the enzyme. Mutagenesis studies have shown that ASP375 is vital to the reaction mechanism and to the hydrolysis reaction (Man *et al.*, 1991). ASP375 in this scheme has to be deprotonated which is in opposition to its proposed state after the condensation reaction. Thus, further insights have to be gained to be able to clarify this conflict.

1.2 CLASSIFICATION OF LIVING ORGANISMS

1.2.1 Woese and his three domains of life

Prior to the seminal ideas of Woese and Fox (1977), living organisms were classed as either eukaryotes (possessing a nucleus) or prokaryotes (lacking a nucleus). Using a phylogenetic analysis based upon ribosomal RNA sequences ,Woese and Fox proposed that living organisms can be split into three basic kingdoms: the eubacteria, the eukaryotes and the archaebacteria. This classification system was updated by Woese *et al.* (1990) proposing the new taxon, a domain, existing above the level of kingdom. Life was now seen as comprising three domains, the Eucarya , the Bacteria and the Archaea, with each domain containing two or more kingdoms.

1.2.2 Lake and his five groups of life

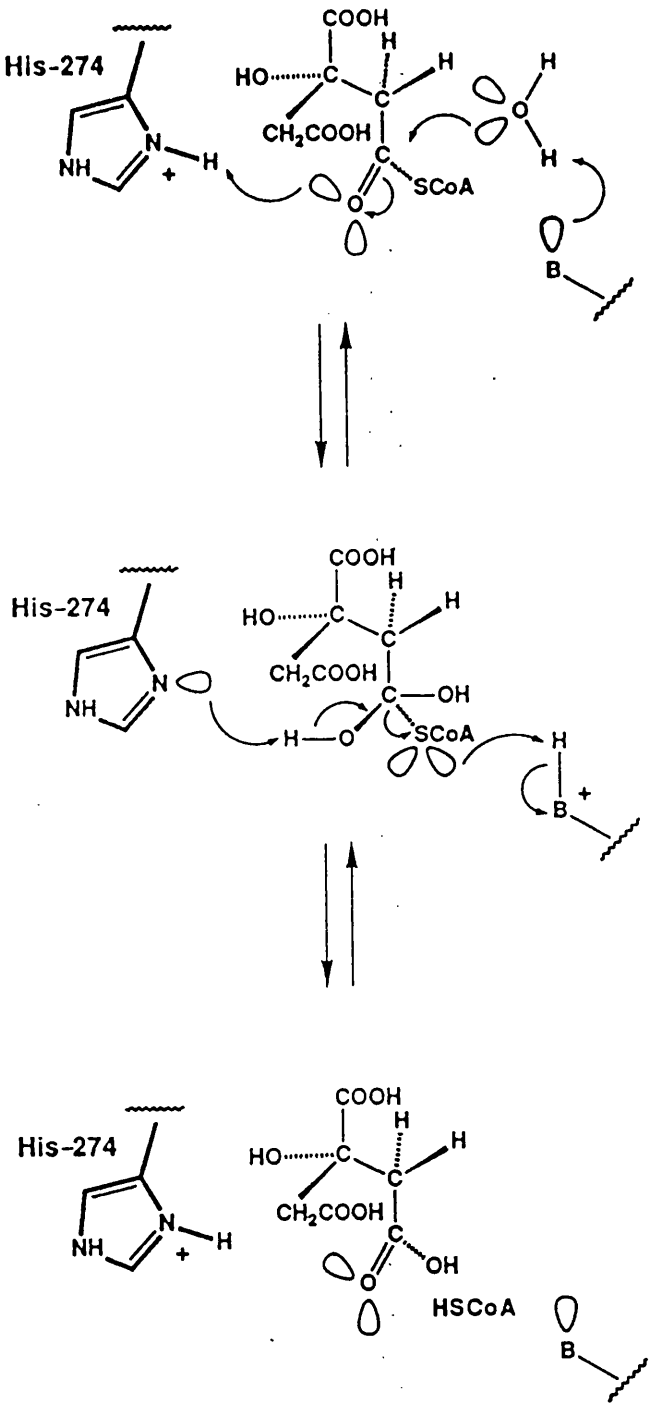
Using refined methods of sequence analysis (see below), Lake (1991) has proposed that living organisms should be divided into five major groups: the Eubacteria, the Halobacteria, the Methanogens, The Eocytes (the sulphur-dependent thermophiles) and the Eukaryotes. Thus, contrary to the view of Woese, the so called Archaea should not be seen as a monophyletic group.

Both these classifications are solely based on phylogenetic analysis which will be discussed in more detail below.

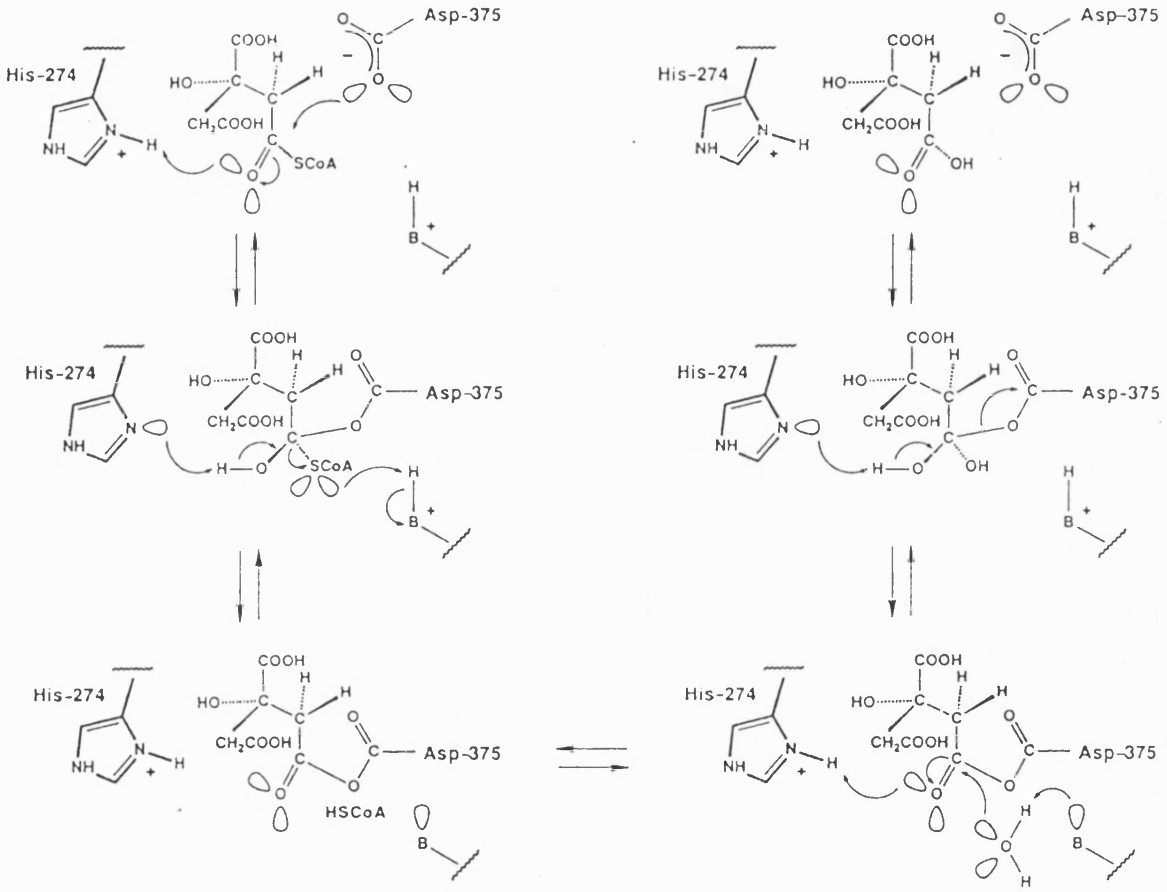
1.3 PHYLOGENETIC ANALYSIS OF LIVING ORGANISMS

1.3.1 The founding father's view – Woese and his rooted three domains

Woese and Fox (1977) were the first to propose an alternative view of phylogenetic classification to the dichotomous one of eukaryotes and prokaryotes. This was made possible by the advance of molecular biological techniques, allowing the sequencing of genes from a diverse range of



Scheme 1.2: Proposed mechanism of hydrolysis in the reaction mechanism of citrate synthase (Remington, 1992). Base-deprotonation of an activated water molecule occurs which attacks the thioester in a concerted reaction involving the deprotonation/protonation of HIS274. The base-abstracted proton is finally transferred to the sulphur of CoA to restore the initial state of the enzyme. Reproduced from Remington (1992).



Scheme 1.3: Alternative mechanism for the hydrolysis stage in the reaction mechanism of citrate synthase, proposed by Alter *et al.* (1990). The thioester is proposed to form an anhydride complex with ASP375, which is then attacked by an activated water molecule. Reproduced from Remington (1992).

organisms. Thus the comparison of gene sequences, in this case 16s rRNA genes, led to the concept of the archaebacteria as a third and distinct kingdom. 16s rRNA is a good molecular chronometer for phylogenetic analysis due to its universality and constancy in function. Upon the elucidation of more and more gene sequences, Woese has developed his ideas and proposed that living organisms should be classified into three domains, the Archaea, the Bacteria and the Eucarya. The current rooted phylogenetic tree showing the relationship between the three domains is shown in Fig. 1.3.

1.3.2 The old pretender's view – Zillig and his fusion event

Although the integrity of the three domains is kept by Zillig *et al.* (1989) the actual rooted phylogenetic tree proposed is different. The tree was derived from studying the gene sequences of RNA polymerases from a wide range of organisms. In comparison to the rooted Woese tree, the tree proposed by Zillig has a central split from a progenote into the archaebacteria and the eubacteria. He proposes that the eukaryotes have arisen from a fusion event between the archaebacteria and the eubacteria (Fig. 1.4) due to the fact that, of the three polymerases found in eukaryotes, pol2 and pol3 are homologous to the archaebacterial counterpart but pol1 is closer to the eubacterial counterpart.

1.3.3 The black sheep's view – Lake and his eocyte tree

Although the archaebacterial tree espoused by Woese is widely hailed as being correct, it has its fierce opponents; the case against is largely led by Lake (1991). Using the same set of data (16s rRNA sequences) but using a different algorithm, he proposes the five grouped eocyte tree - Fig. 1.5. There is a deep cleft near the root of the tree dividing living organisms into

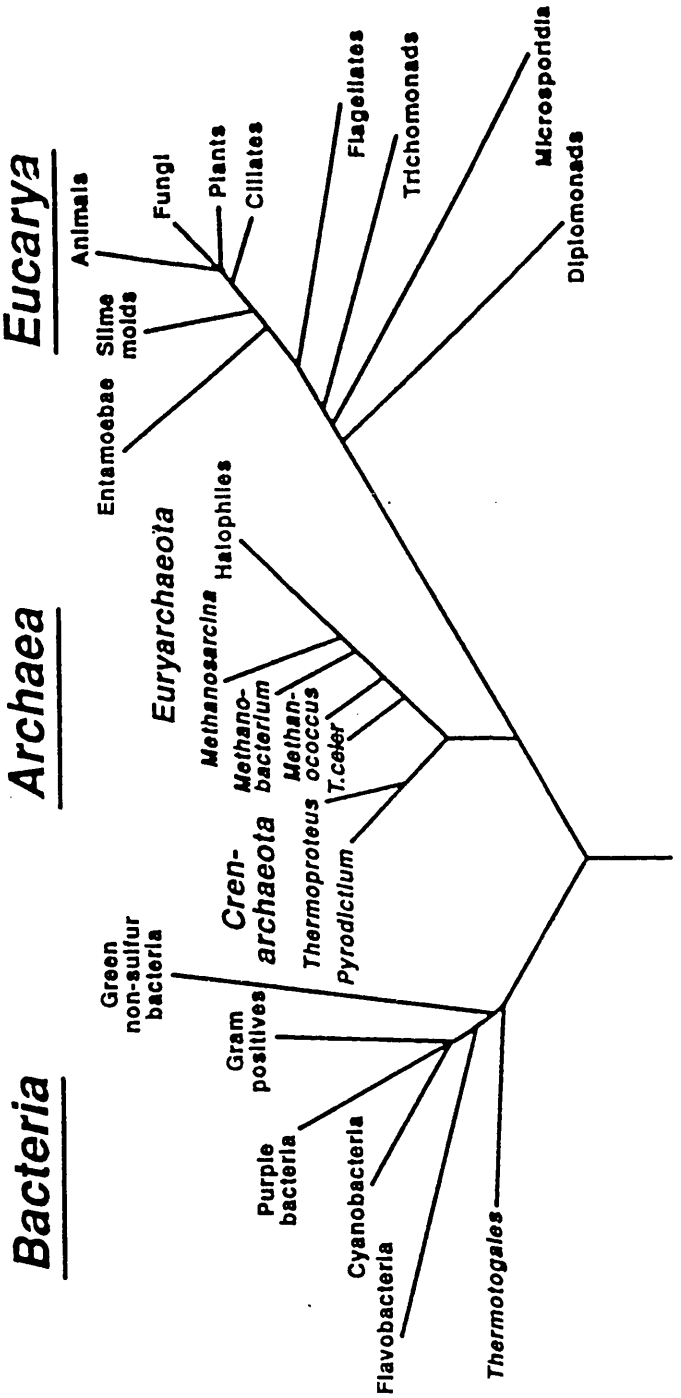


Figure 1.3: Rooted phylogenetic tree, as proposed by Woese, displaying the three domains of life based upon 16S rRNA sequences. The root of the tree was inferred from ancestral gene duplications. Reproduced from Olsen and Woese (1993).

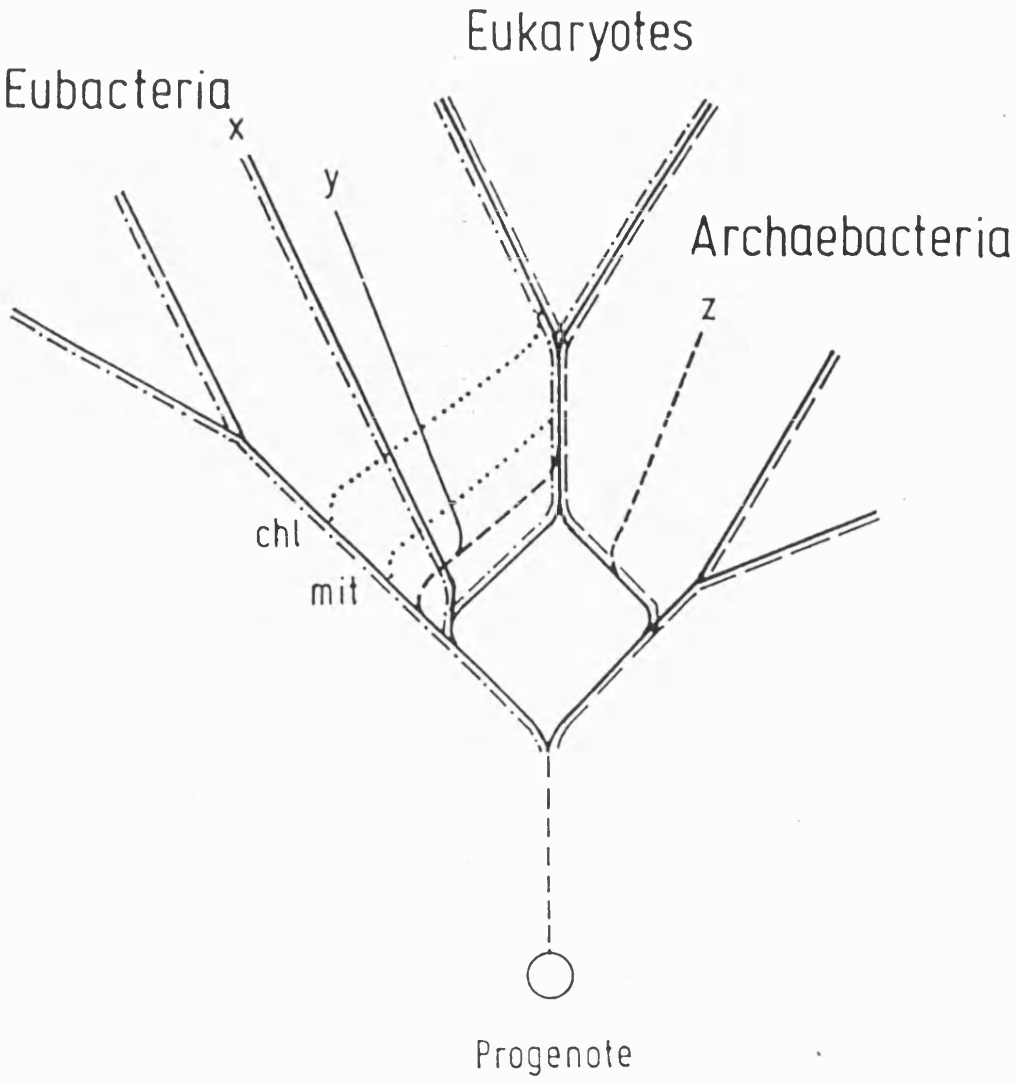


Figure 1.4: Schematic representation of the fusion event as an explanation for the evolution of organisms proposed by Zillig *et al.* (1989). The dotted lines represent the acquisition of mitochondria and chloroplasts by the eukaryotes. The intermediary lines X, Y and Z show how for certain molecules certain eubacterial and archaebacterial lineages could branch from the eukaryotic lineages. The dot-dash line represents molecules (eg. pol1) that were inherited from an eubacterial ancestor and the interrupted line represents molecules (eg. pol2 and pol3) that were inherited from an archaebacterial ancestor. Reproduced from Zillig *et al.* (1989).

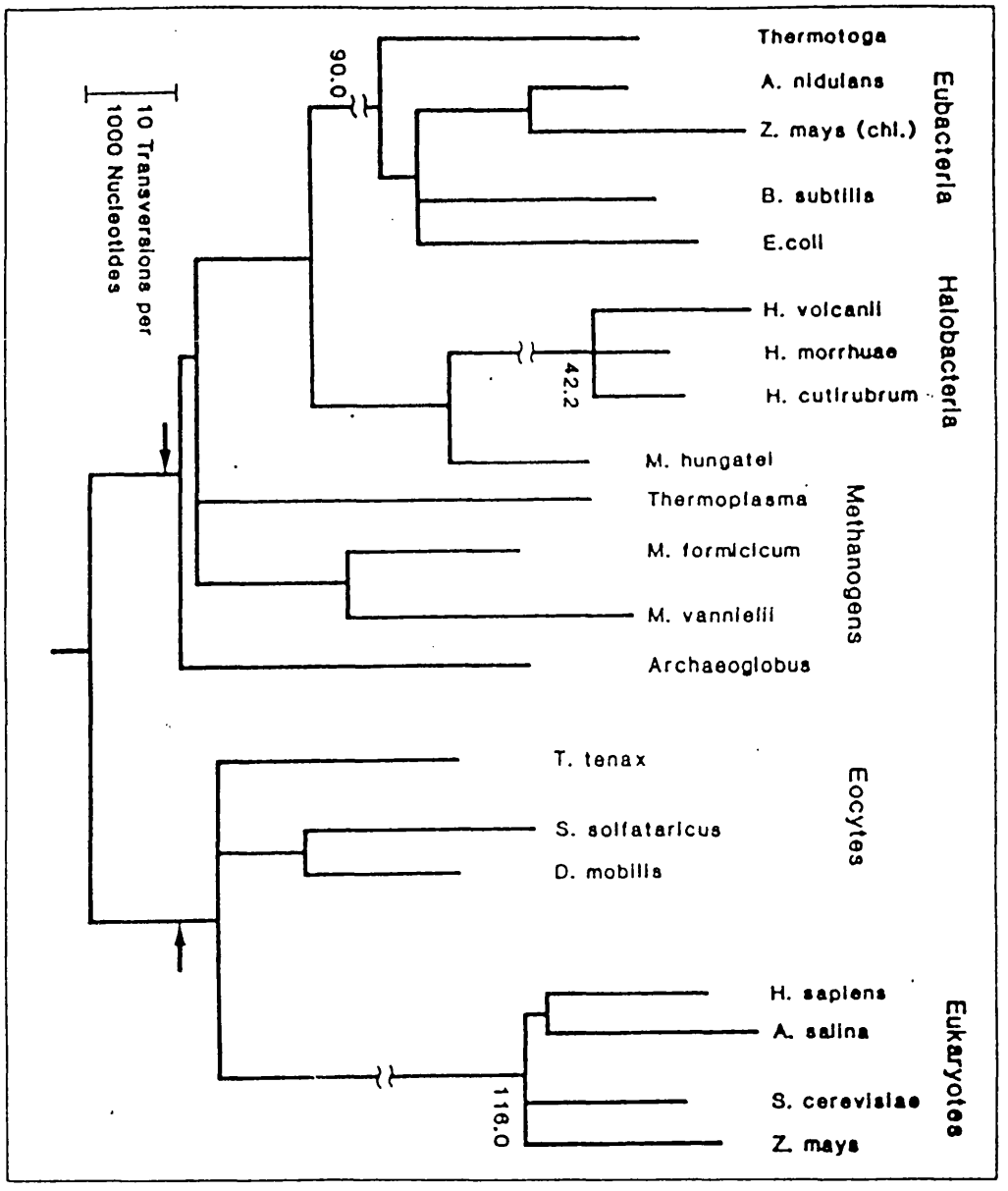


Figure 1.5: The rooted evolutionary tree as proposed by Lake, showing a deep division which separates the tree into two monophyletic groups. The rapidly evolving Eubacterial, eukaryotic and halobacterial branches have been shortened to fit the figure. Reproduced from Lake (1991).

two monophyletic superkingdoms, the parkaryotes and the karyotes. The former comprises the eubacterial, the halobacterial and the methanogenic groups and the latter the eocyte (the hyperthermophiles) and eukaryotic groups.

1.3.4 Why the confusion?

The majority of controversy over phylogenetic analysis of living organisms has been between Woese and his many followers and Lake. Even though both camps were analysing the same set of sequence data, the differing end results have provoked wide debate about the validity of the algorithms used and the results drawn from them. Both the Lake and Woese algorithms are mathematically correct but it is the assumptions concerning the events of microevolution that differ. Neither of the two camps use the whole rRNA sequence and actually select different regions for subsequent analysis. Lake has argued that Woese results are dogged by large unequal rate effects (arising from the fact that the rates of change of a nucleotide are different in adjacent branches of a phylogenetic tree) which produces unreliable analyses. Woese reposted. He insists that Lake's analysis is affected by an unspecified systematic error (Olsen and Woese, 1989).

1.3.5 Why choose the Woese convention?

The most widely used and accepted classification scheme is that of Woese and therefore, the wide use (including this thesis) of the associated nomenclature (that of the Eucarya, the Bacteria and the Archaea). Other than the use of rRNA and proteins as phylogenetic markers, evidence supporting the existence of the Archaea as a distinct domain comes from studies revealing unique features of the physiology and biochemistry of the Archaea (Kandler, 1993).

1.4 THE ARCHAEA

The Archaea comprise three phenotypically distinct groups: the methanogens, the sulphur-dependent thermophiles and the halophiles (Woese *et al.* 1978). The methanogens are obligate anaerobes that produce methane from hydrogen and carbon dioxide, or other simple C_1 compounds (reviewed by Jones *et al.* 1987). The sulphur-dependent thermophiles grow at high temperatures (up to 113 $^{\circ}C$), usually anaerobically, and utilise sulphur as an electron acceptor (reviewed by Stetter and Zillig 1985). The halophiles grow in high salt conditions and maintain a high level of internal salt (up to 3.5M KCl) (reviewed by Kushner, 1985).

On the basis on rRNA and protein phylogenetic analysis (the Woese variety) the Archaea can be split into two basic groups: the euryarchaeota (halophiles/methanogens) and the crenarchaeota (sulphur-dependent thermophiles). *Tp.acidophilum* has been shown to belong to the former group although it is a thermophile, but recent studies suggest that it belongs to a unique branch within the Archaea (see section 1.5.1).

1.4.1 Biotechnological potential of the Archaea

The adaptations necessary for the Archaea to exist (and thrive) in extreme environments will yield a wealth of information concerning all aspects of stability. The insights into how proteins, for example, are stabilised and the discovery/isolation of highly stable enzymes from these organisms may be extremely valuable to biotechnological industries (Hough and Danson, 1989).

1.5 *THERMOPLASMA ACIDOPHILUM*

1.5.1 Classification of *Tp.acidophilum*

Tp.acidophilum was first isolated from burning coal-refuse piles by Darland *et al.* (1970) and has since been discovered in naturally occurring hot springs and sulfataric fields (Segerer *et al.*, 1988). The organism is an aerobic heterotroph living at temperatures between 50–64°C and between pH 0.5–3.0.

Tp.acidophilum is a thermoacidophilic Archaeon. Previous studies (Yang *et al.*, 1985) on 16S rRNA sequence comparison led to the proposal that the organism belonged to the methanogen-halophile branch of the then archaebacteria, later reclassified the euryarchaeota (see section 1.4). However, phenotypically it shows a closer resemblance to the sulphur-dependent thermophile branch, the crenarchaeota. A recent study of the primary and secondary structures of 23S rRNA by Ree *et al.* (1993) shows that 23S rRNA from *Tp.acidophilum* shares secondary structural characteristics with both the euryarchaeota and the crenarchaeota. From these results and from a sequence-derived phylogenetic tree, they propose that *Tp.acidophilum* stems from an ancient, and possibly unique, divergence within the archaeal domain - see Fig. 1.6.

1.5.2 Environmental adaptation of *Tp.acidophilum*

Tp.acidophilum grows optimally at moderately high temperatures and low pH values and therefore must possess adaptive mechanisms to withstand these extremes. Although *Tp.acidophilum* grows at a low pH (0.5–3.0) it maintains an internal cytoplasmic pH value of approximately pH6. A H^+-translocating ATPase is known to regulate the internal pH of other acidophiles, but this enzyme is not present in *Tp.acidophilum.*

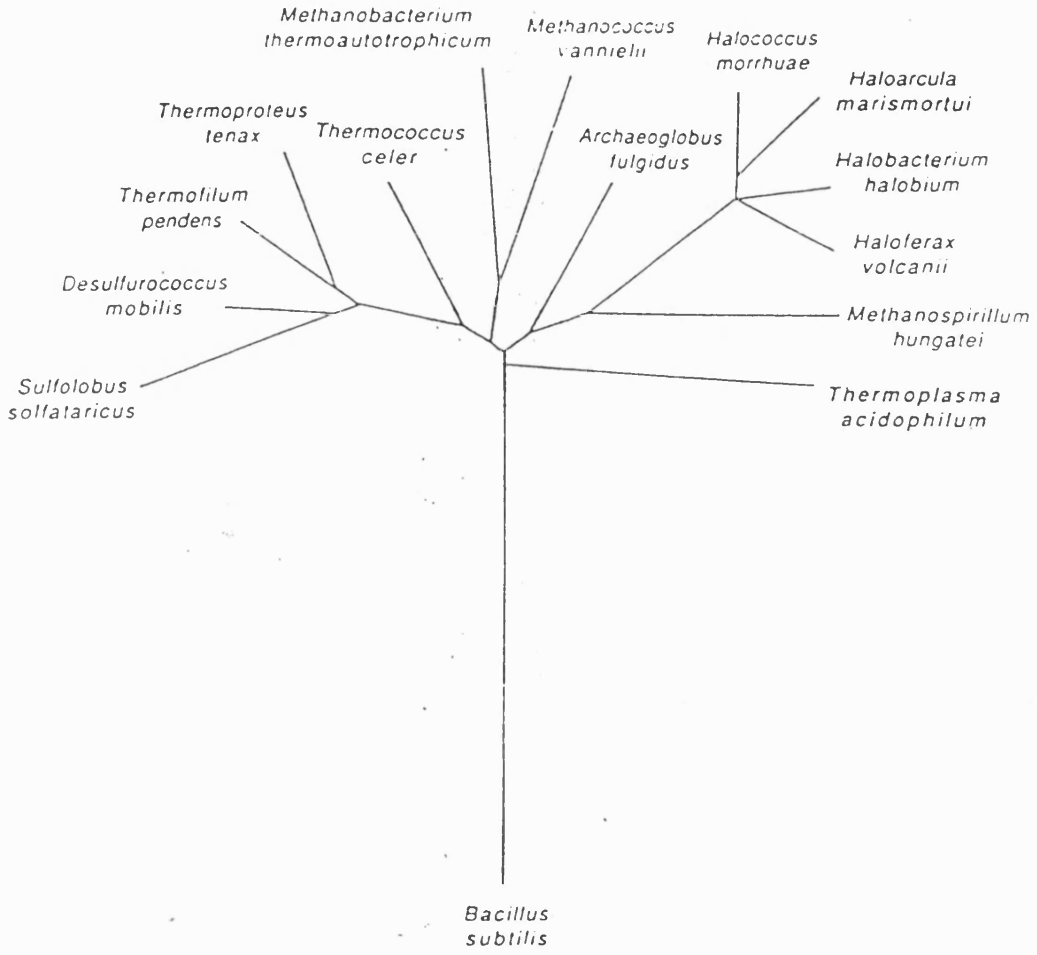


Figure 1.6: Phylogenetic tree of the Archaea based on 23S rRNA sequence analysis proposed by Ree *et al.* (1993), with the bacterium *Bacillus subtilis* serving as an outgroup. *Tp.acidophilum* is shown to stem from an ancient divergence within the Archaeal domain. Reproduced from Ree *et al.* (1993).

Searcy (1986) proposed that this absence is probably due to the extreme environmental living conditions of *Tp.acidophilum* and that the organism possesses an electron transport chain for maintenance of its internal pH.

Tp.acidophilum, along with all other Archaea and some bacterial thermophiles, possesses isopranyl glycerol ether-linked lipids in contrast to fatty acid glycerol ester-linked lipids in all other living organisms. All archaeal glycerol ether lipids also contain the unusual 2,3-sn-glycerol (Gambacorta *et al.*, 1993). The presence of these lipids serve as a marker for archaeal organisms, and may reflect an essential adaptation for extremophilic growth.

Tp.acidophilum DNA has an average G–C base composition of 46% (Christiansen *et al.*, 1975; Searcy and Doyle, 1975). High G–C content is known to cause increased relative thermal stability of DNA, but in this instance some factor extrinsic to the DNA must be necessary for the thermal stabilization. A histone-like protein (HTa) has been identified and purified from *Tp.acidophilum* and has been shown to bind specifically to DNA (Searcy, 1975; DeLange *et al.*, 1981), causing a stabilization of the DNA against thermal denaturation (Stein and Searcy, 1978).

Amino acid sequences of a number of *Tp.acidophilum* proteins are known, either from amino acid sequencing (HTa protein (Delange *et al,* 1981) and ferredoxin (Wakabayashi *et al.*, 1983)) or from DNA sequences (for example, citrate synthase (Sutherland *et al.*, 1990), glucose dehydrogenase (Bright *et al.*, 1993) and elongation factor-2 (Pechmann *et al.*, 1991), but sequence comparison studies tend to be poor in identifying features responsible for thermal stabilization of proteins. 3-dimensional structures of proteins are much more valuable tools for gaining insights into the thermal stabilization of proteins but to date only one enzyme (glucose dehydrogenase (J.John personal communication) from *Tp.acidophilum* and only one archaeal pro-

tein (Rubredoxin from *Pyrococcus furiosus* (Day *et al.*, 1992) have had their 3-dimensional structures elucidated.

1.5.3 *Tp.acidophilum* citrate synthase

The kinetic parameters of *Tp.acidophilum* citrate synthase in relation to other archaeal citrate synthases were studied by Danson *et al.* (1985). Smith *et al.* (1987) purified *Tp.acidophilum* citrate synthase and showed the enzyme to be present as an active dimer with subunit M_r 43000 (+/- 2000). The gene for the enzyme was cloned and sequenced (Sutherland *et al.*, 1990) and subsequently overexpressed in *E.coli* (Sutherland *et al.*, 1991). *Tp.acidophilum* citrate synthase was found to be approximately 50 residues shorter than all other citrate synthases known at the time, and from sequence alignment studies this was attributed to an N-terminal extension in the other citrate synthases. More recent studies have revealed other archaeal citrate synthases (*Pyrococcus furiosus* (Muir *et al.*, 1993), *Haloferax volcanii* (James *et al.*, 1991) and *Sulfolobus acidocaldarius* (Lill *et al.*, 1992) to be N-terminally stunted, a feature also present in citrate synthase from a thermotolerant *Bacillus* species (Schendel *et al.*, 1992). *Tp.acidophilum* citrate synthase shows a mean identity of 28% to bacterial citrate synthases and a mean sequence identity of 20% to eucaryal citrate synthases. Despite this low sequence identity, sequence alignment studies indicate conservation of 8 out of 11 residues vital for catalytic activity in pig heart citrate synthase (Sutherland *et al.*, 1990), suggesting a similar mechanism of catalysis and 3-dimensional structure for both the eucaryal and archaeal enzymes.

1.6 PROTEIN STABILITY

1.6.1 Introduction

A fundamental, and as yet not understood, question concerning protein structure is how proteins of the same function from different sources can have large differences in their stability despite possessing a very similar folded conformation. The earliest life on earth was thought to be thermophilic in nature and proteins from thermophilic sources, especially archaeal ones, may be seen as ancestral forms. Changes seen in their mesophilic counterparts may then account for differences in stability. Therefore, structural studies on the homologous proteins from sources living in a wide range of habitats should yield information as to why some proteins are more stable than others.

Protein stability, per se, can be seen as the difference in stability between the folded and unfolded states of the protein. The conformation of the folded native state can be seen as the global energy minimum, and the difference in free energy between the folded and unfolded states as the conformational stability of the protein. Globular proteins have been shown to have a conformational stability of 5-15 $Kcalmol^{-1}$. This is a relatively small value and must reflect an energetic balance between the need for stability and the need for protein flexibility and eventual turnover. The forces which govern this conformational stability are acting so as to reduce the free energy of the folded state relative to the unfolded state.

There are 6 basic forces that affect the free energy of the folded state relative to the unfolded state: conformational entropy, disulphide bonds, hydrogen bonds, ionic interactions, hydrophobic interactions and Van der Waals forces.

Thermal stability, and hence protein stability, has been studied in one of

three basic ways - primary sequence studies, elucidation of the 3-dimensional structure of proteins and site-directed mutagenesis studies. Although the first two types of study have relied on comparisons of proteins of differing stability, the majority of mutagenesis studies have concentrated on specific, well-characterised mesophilic proteins in order to increase the stability of the mutant with respect to the wild type protein. Each of these three ways will be discussed in more detail below.

1.6.2 Primary sequence comparisons

The majority of the work in this area has concentrated on amino acid replacements from mesophile to thermophile where a tertiary sequence was known for one of the members of the comparison which allows structural positioning of the observed changes. Menendez-Arias and Argos (1989) updated and refined the work of Argos *et al.* (1979) due to the increase in available mesophilic and thermophilic sequences. From this study they were able to reveal general trends in mesophile to thermophile amino acid replacements, resulting in a top ten of most common substitutions, with LYS->ARG and SER->ALA being the first and second in this list. The former correlates well with an earlier study of Merkler *et al.* (1981) who reported that the ARG/(ARG+LYS) ratio of proteins correlates with the optimum growth temperature of the organisms from which the proteins were isolated. The top ten changes were proposed to increase the hydrophobicity and decrease the flexibility of the thermophilic protein. The majority of the changes were localised in α-helices and/or domain/subunit interfaces. Although specific amino acid substitutions were identified, they proposed that it was unlikely that a single replacement would cause a large increase in thermostability.

Ikai (1980) carried out a statistical analysis of protein sequences by

investigating the relative volume of a protein occupied by aliphatic side chains (ie. ALA, VAL, ILE and LEU) - the aliphatic index. Proteins from thermophilic hosts showed a significantly higher index value than those from mesophilic hosts.

1.6.3 Structural analysis of homologous mesophilic and thermophilic proteins

The elucidation of homologous structures of varying stability has allowed direct structural comparisons, and therefore accurate positioning of specific changes, and comparisons of specific interactions. A relatively small number of thermophilic proteins have had their crystal structures elucidated and a few specific examples will be discussed below. In all known cases to date the mesophilic and thermophilic proteins have had very similar overall conformations allowing accurate pinpointing of any minor structural differences which may account for their differing stability.

The first structural investigation into thermostability was preformed by Perutz and Raidt (1975) who compared mesophilic and thermophilic ferredoxins. They reported that the presence of a few extra internal salt bridges were the most likely causative agent for the increased stability.

Ishikawa *et al.* (1993) elucidated the structure of the thermophilic Ribonuclease H (RNaseH) from *Thermus thermophilus* allowing direct comparisons to its mesophilic counterpart from *E.coli.* A large cluster of 8 aromatic residues are seen in the former, a known stabilizing feature in proteins (Burley and Petsko, 1985). An increase from 9 to 14 salt bridges was also observed in the thermophile compared to the mesophile, 4 of which are intra-helical salt bridges that serve to cancel out the macro-dipole of the respective helix. Finally, they observed a substitution of a lysine in a left-handed helical conformation in the mesophile with a glycine residue in

the same conformation in the thermophile. This was proposed to decrease the steric hindrance between the β-carbon atom and the carbonyl oxygen atom within the same residue.

Davies *et al.* (1993) compared the structure of phosphoglycerate kinase from the thermophile *Bacillus steareothermophilus* with the same enzyme from yeast. They observed an additional 15 electrostatic interactions in the thermophile, 10 of which are concentrated in the N-terminal domain of the protein, a region which is known to be the least stable part of the mesophilic protein.

The above three examples describe putative stabilizing interactions in thermophilic monomeric proteins. However, a large number of proteins exist functionally as oligomers and therefore they may reveal specific interactions that are necessary to maintain the integrity of the oligomer.

Kelly *et al.* (1993) elucidated the structure of malate dehydrogenase from *Thermus flavus*, an enzyme stable up to $90°C$. On comparison to the homologous (55% sequence identity) enzyme from *E.coli*, they found good correlation with the work of Menendez-Arias and Argos (see above) with respect to both sequence changes and the position of these changes. There was a marked increase in alanine content of the thermophile, with the majority of the changes occurring in solvent-accessible α-helices. 4 extra ion pairs were also present, three in the subunit-subunit interface and one in the domain interface.

A limited amount of stability analysis was carried out on the thermostable enzymes isopropylmalate dehydrogenase (Imada *et al.*, 1991) and glyceraldehyde-3-phosphate dehydrogenase (Walker *et al.*, 1980) from *Thermus thermophilus* and *Bacillus steareothermophilus*, respectively. In both of these enzymes an increase in hydrophobic interactions were observed in

inter-subunit contacts.

There are only a very limited number of homologous mesophilic and thermophilic protein structures known and as yet no obvious trend of stabilizing features has been observed. This may be due to the relatively low sequence identity between homologues with neutral amino acid replacements resulting in a masking of stabilizing features, or alternatively may be due to different proteins having different methods of stabilization. This conflict may be resolved with the elucidation of the crystal structure of more thermophilic proteins. An alternative approach is the use of site-specific mutagenesis experiments in order to probe the contribution of specific amino acids to global protein stability.

1.6.4 Mutagenesis experiments

The majority of the work in this field has concentrated on two proteins - Barnase (the ribonuclease from *Bacillus amyloliquefaciens*) from the Fersht group and lysozyme (from the bacteriophage T4) from the group of Matthews. Nevertheless, many experiments have been carried out that have been rationally designed through homologous structure studies.

1.6.4.1 Structural context

The most important consideration in the design of mutagenesis experiments to test theories concerning protein stability is the structural context of the proposed change. For example, engineered salt bridges on the surface of T4 lysozyme produce no significant change in stability (Dao-Pin *et al.* ,1991) and no change in the structure of the mutant protein. The added interaction on the surface is probably off-set by entropic cost of restricting the motion of the ion pairs. On the other hand, the disruption of a buried electrostatic interaction in T4 lysozyme (HIS31-ASP70) created a mutant

protein of decreased stability (Anderson *et al.*, 1990). This concept of structural context was demonstrated dramatically by Rennell *et al.* (1991) who substituted 163 of the 164 residues of T4 lysozyme with 13 different amino acids resulting in 2015 different single amino acids substitutions. Only 173 of these mutations resulted in a significant destabilisation and they were almost all localised to regions of the protein that have low mobility.

1.6.4.2 Hydrogen bonds

The contribution of hydrogen bonds to the stability of the folded state has been investigated by mutation of known H-bond acceptors and donors within a protein of known structure. Mutation of a charged H-bond acceptor/donor results in a larger decrease in stability than that of an uncharged one in tyrosyl-tRNA synthetase (Fersht *et al.*, 1985). Shirley *et al.* (1992) have demonstrated that H-bonding in RNase T1 contributes a large extent to the protein's conformational stability.

1.6.4.3 Hydrophobic interactions

The burial of hydrophobic residues is believed to be the major driving force of protein folding and therefore a major contributor to stabilization of the folded state. Studies on this force have concentrated on cavity forming mutants by creating small deletions in side chains without changing their geometry, eg. ILE->VAL, ALA->GLY and LEU->ALA. Typically 1.5 $KJmol^{-1}$ of stability is lost per methylene group deleted (Kellis *et al.*, 1988). Crystal structures of the mutants usually reveal a rearrangement of the structure to compensate for the loss of missing interactions. Eriksonn *et al.* (1992) solved the crystal structure of several LEU->ALA cavity forming mutants in T4 lysozyme. In every mutant a cavity was present, but in some instances large shifts of up to 1Å in some parts of the protein were observed

as the cavity vacated by the LEU residue was filled by other atoms. In other cases, the mutant structure did not differ significantly from the wild type and therefore larger cavities were observed than in the structurally-rearranged mutant. The size of the cavity was shown to correlate in a linear fashion with the loss of stability observed. The generated cavity is less likely to collapse if the hydrophobic packing density is high in the wild type because of stronger van der Waals interactions with the methylene group of interest and the surrounding side chains. Removal of more than one methylene group leads to a cumulative decrease in stability and therefore, in the folded state, the hydrophobic interactions probably stabilize the protein cumulatively.

Destabilization of the wild type protein by cavity forming mutations have been reasonably successful, but stabilization by improvement of the packing of the hydrophobic core has had only limited success. Eijsink *et al.* (1992) created two mutants (LEU–>TRP and MET–>TRP) in the hydrophobic core of the neutral protease from *Bacillus subtilis*, but only increased the thermal stability by a maximum of $0.4°C$. This is probably due to increased internal strain caused by the bulkier side-chains.

1.6.4.4 Conformational entropy

Investigations into the contribution of conformational entropy to protein stability have been carried out by Matthews *et al.* (1987). This force is the major opposing force to protein folding/stability. Two substitutions in T4 lysozyme were made to increase the bulk of specific residues (GLY–>ALA and ALA–>PRO) without causing any bad contacts, and hence reduce the conformational flexibility of the protein. These two changes resulted in only a slight increase in stability.

1.6.4.5 Disulphide bonds

The engineering of disulphide bonds into specific parts of the protein has had variable success in increasing stability. 4 different S-S bonds were mutated into T4 lysozyme which increased the thermal stability between 5-10°C (Matsumura *et al.*, 1989). A larger increase was probably not seen due to additional strain energy produced by the formation of the S-S bond within the protein. This strain can be reduced if the S-S bond is engineered into flexible parts of the protein and a large loop is formed by the new bond. Gokhale *et al.* (1994) increased the conformational stability of thymidylate synthase from the mesophile *Lactobacillus casei* by nearly 40°C by creating a mutant with 2 S-S bonds across the dimer interface. The mutant protein was optimally active at 55°C and still retained some activity at 65°C.

1.6.4.6 α-Helical stability

A large amount of mutational studies has concentrated on the factors that contribute to α-helical stability. Horovitz *et al.* (1992) created 19 different mutations of an internal ALA residue in helix 32 in Barnase, equivalent to all other amino acids, to generate a helix-forming propensity scale for amino acids. In agreement with studies on synthetic peptides (O'Neil and Degrado, 1990), alanine was found to be the most stabilizing with proline the least. Although many studies correlate with these findings the surrounding residues, the position of the amino acid in the helix and the position of the helix itself within the protein may contribute to specific helix stability. The helix propensity scale can be rationalized energetically in one of three ways; the burial of hydrophobic surface and van der Waals surface, differences in the conformational entropy of each amino acid in the helix relative to the unfolded state and finally, differing solvation effects.

One of the most consistent ways to increase the stability of a protein by

site-directed mutagenesis is the mutation of the helix end residues. Helices have an inherent macro-dipole along their length with a positive charge at the N-terminus and a negative charge at the C-terminus. Mutagenesis studies (for example, Serrano *et al.*, 1992) replacing the N_{cap} residue with a negatively-charged one (eg GLU or ASP) or a neutral polar residue (SER or THR) or the C_{cap} residue with a positively charged residue (eg LYS, ARG or HIS) tends to stabilize the helix and the protein as a whole.

1.6.4.7 Specific examples

The majority of the above mutations have caused only small rises in the melting temperature of the protein, but there have been a few specific examples where dramatic increases in stability have been produced. Kotik and Zuber (1993) mutated two residues (SER->ALA and THR->ALA) in α-B helix which lies in the subunit interface in the tetrameric lactate dehydrogenase (LDH) from the mesophilic bacterium *Bacillus megaterium.* The mutant protein exhibited a $20°C$ rise in thermostability with respect to the wild type protein and also increased the rigidity of the whole enzyme. The mutations were rationally designed by sequence comparisons with the closely-related LDH from the thermophilic bacterium *Bacillus steareothermophilus.*

Mutagenesis studies on *E.coli* RNaseH confirmed findings from structural comparisons between the mesophile and thermophile. A mutant was created in which the lysine in the left-handed helical conformation (see above) was replaced by the glycine residue seen in the thermophile (Kimura *et al.*, 1992a). A $6.8°C$ rise in the thermostability of the mutant was seen with respect to the wild type. A mutant of the mesophilic protein was also made which had a series of substitutions in its equivalent region to the aromatic cluster region of the thermophile, in order to make the mutant partly

resemble the thermophile. An $8.2°C$ rise in thermostability was observed for the mutant with respect to the wild type (Kimura *et al.*, 1992b).

1.6.5 Concluding remarks

A vast wealth of information has been gained on specific factors affecting protein stability, primarily from the seemingly countless number of mutagenesis experiments. However, the ability to design on a rational basis mutants of significantly increased stability, seems still to be some way in the future. The majority of the mutants come from a small range of small proteins and it is clear that structural information is needed on a wider range of thermostable proteins. Eventually, general trends governing protein stability and enhanced stability may be forthcoming, but what is already clear is that different proteins may be stabilized in different ways and the power of cumulative interactions on the protein as a whole is vitally important. Nevertheless, confident predictions of factors governing the stability of a specific protein will be able to be made only if the crystal structure of that protein is known.

1.7 BACKGROUND AND AIMS

Citrate synthase was chosen as a model enzyme system from which information could be gained concerning both the archaeal features of proteins and the features that confer thermostability. The gene for citrate synthase from the thermoacidophilic Archaeon *Thermoplasma acidophilum* had previously been cloned and sequenced (Sutherland *et al.*, 1990) and expressed in *E.coli* (Sutherland *et al.*, 1991). The aims of this work were initially, to gain sufficient quantities of pure protein for wide spread crystallization trials. Once suitable crystals had been grown, structural elucidation was hoped to proceed using the known crystal structure of the mesophilic pig heart citrate synthase as a search model in molecular replacement techniques. If the structural elucidation of the thermophilic citrate synthase proved to be successful, detailed structural comparisons could be carried out between pig heart and *Tp.acidophilum* citrate synthases in order to highlight structural features conferring thermal stability on the archaeal enzyme.

Chapter 2 : Experimental Techniques and Programs

2.1 MATERIALS

2.1.1 Enzymes,reagents and other materials

5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB); oxaloacetate and ampicillin were from Sigma, Poole, UK. Tryptone and yeast extract were from Difco, Michigan, USA. CoenzymeA was from Boehringer Mannheim, Germany. Microsep microconcentrators were from Flowgen, UK. PBE118, pharmalyte, protein molecular weight standards and Phast gels were from Pharmacia-LKB, UK.

All other chemicals were from BDH, Fisons and Sigma (all UK). All solvents (standard laboratory grade) were from BDH, UK and Fisons, Loughborough, UK.

2.2 METHODS

2.2.1 Assay for citrate synthase activity

Samples were assayed spectrophotometrically for *Thermoplasma acidophilum* citrate synthase activity at $55°C$ according to the method of Srere *et al.* (1963). The assay mixture contained 20mM Tris-HCl (pH8.0), 1mM EDTA, 0.2mM acetyl CoA, 0.2mM oxaloacetate and 0.1mM DTNB. Acetyl CoA was prepared using the method of Stadtman (1957). The assay was started by adding enzyme and the increase in absorbance monitored at 412nm (E_{412} 13600 $mol^{-1}cm^{-1}$).

2.2.2 Bradford protein concentration determination

900μl Bradford assay mixture (0.01%(w/v) Coomassie Blue G.250, 5% ethanol, 8.5% H_3PO_4) was added to 100μl protein sample. The sample

was measured spectrophotometrically at 595nm (Bradford, 1976). Protein concentrations were read off a Bovine serum albumin calibration curve of a range of 0-25μg.

2.2.3 Production of *Thermoplasma acidophilum* citrate synthase in *Escherichia coli*

10ml cultures of *E.coli*, containing the plasmid (pCSEH19) which carried the *Tp.acidophilum* citrate synthase gene, was grown for 17hr at 32°*C* in 2xYT medium containing ampicillin (25μg/ml). An equal amount of ampicillin was added and the culture incubated for another hour. This culture was then used to subinoculate 1l 2xYT medium containing ampicillin (25μg/ml) and the above incubation carried out.

2.2.4 Preparation of cell extracts from *E.coli*

Cells from 1l cultures were harvested by centrifugation (5000g at 4°*C* for 10min), resuspended 20mM Tris-HCl (pH8.0) and 2mM EDTA and sonicated with a 20mm probe at 180 Watts for 2min. Cell debris was removed by centrifugation (12000g at 4°*C* for 20min).

2.2.5 Heat purification

Cell free extracts of *E.coli* were heated in a glass tube for 15min at 65°*C*. Denatured protein was then removed by centrifugation (10000g at 4°*C* for 20min)

2.2.6 Gel filtration

The supernatant from the heat treatment was then passed down a FPLC Superdex G-200 gel filtration column (Pharmacia-LKB) pre-equilibrated with 25mM triethylamine (pH11.0) or a suitable crystallization buffer.

2.2.7 Chromatofocussing

Chromatofocussing was carried out on a column packed with Polybuffer exchanger118 (PBE118). The fractions from the gel filtration with *Tp.acidophilum* citrate synthase activity were loaded onto the PBE118 column pre-equilibrated with 25mM triethylamine (pH11.0) and eluted with pharmalyte (pH8.0).

2.2.8 Matrex Red GelA affinity chromatography

The heat treated extract containing *Tp.acidophilum* citrate synthase activity was loaded onto a 20ml column packed with Matrex Red GelA pre-equilibrated with 20mM Tris-HCl and 1mM EDTA buffer. Elution buffer contained 20mM Tris-HCl, 1mM EDTA, 5mM oxaloacetate and 1mM CoenzymeA.

2.2.9 Crystallization trials

The hanging drop vapour diffusion method was used for all the trials (McPherson, 1990). The precipitants used included PEG3350 (3%-24%), PEG6000 (2.5%-11%), PEG800 (7%-9%) and $(NH_4)_2SO_4$ (20%-70%). The pH range investigated was 3.5-9.5. The matrices were repeated at 25°C and at 4°C.

2.2.10 Thermostability assay

A known volume of 20mM Tris-HCl (pH8.0) and 1mM EDTA buffer was equilibrated to a known temperature, measured via a temperature probe. A small volume of extract was added giving a known dilution. After a known time a 10μl aliquot of the mixture was assayed for citrate synthase activity and this was repeated over a time course.

2.2.11 Circular dichroism spectra

All spectrum were recorded on a JASCO J-600 spectropolarimeter using a 0.1cm pathlength cuvette. A protein concentration of 48μg/ml in 10mM Tris-HCl (pH 8.0) buffer was used throughout.

2.3 PROGRAMS USED

Data Collection and Analysis

XDS[1]: Automatic processing of Xentronics data frame to produce list of scaled hkl intensities. Described in detail in Chapter 5. Version2 by Kabsch (1988).

XSCALE[1]: Applies empirical absorption coefficient to individual intensities. Allows scaling and merging of individual data sets. Version (October 1991) by Kabsch.

XX[1]: Modified version (G.L. Taylor) of XSCALE allowing preservation of intensities if Friedel pairs (I+, I-), for use in heavy atom derivative data analysis.

X2L[1]: Modified version (G.L. Taylor) of OXMAKE (CCP4) to convert the scaled list of intensities to structure factors.

Molecular Replacement

MERLOT[2,3]: Integrated and rationalised package of previously existing programs. Version 2.3 by P.M.D. Fitzgerald (1988).

POLARRFN[2]: W. Kabsch, CCP4 Program suite. A fast rotation function program which produces sections of constant rotation angle kappa for different axis directions defined by omega and phi. The sections can be plotted as stereograms.

X-PLOR[3]: Version 2.1 devised by A.T. Brünger (1990).

AMORE[3] Version 2.1.4 devised by J. Navaza (1992).

Heavy atom isomorphous replacement analysis

MTZUTILS[2]: E. Dodson and H. Terry, CCP4 Program suite. Creates new re-arranged or edited MTZ reflection data file from one or two existing files.

LOCAL[2]: CCP4 program Suite. Anisotropic scaling of heavy atom derivative data to the native data followed by local scaling.

FFT[2]: L.F. TenEyck, CCP4 Program suite. Allows calculation of Fouriers, difference Fouriers, Pattersons and difference Pattersons.

PLUTO[2]: S. Motherwell (Modified by P. Evans), CCP4 Program suite. Allows plotting of contour maps.

VECSUM[2]: CCP4 Program suite. Deconvolutes a Patterson function and allows semi-automatic interpretation of a difference Patterson map.

VECREF[2]: I. Tickle, CCP4 Program suite. Vector space refinement of heavy atom sites in isomorphous derivatives.

MLPHARE[2]: E.J. Dodson, CCP4 Program suite. Refines heavy atom positions and generates phase information.

Refinement and Analysis

X-PLOR[3]: Version 2.1 devised by A.T. Brünger (1990).

ENVIRONMENTS[2]: Creates 3D-1D score and profile plot of new structure. Luthy *et al.*, 1992

PROCHECK[2]: Analysis of stereochemical quality of protein structure

coordinates. Version 2.1 by Laskowski *et al.*, 1993.

SHP[2]: D. I. Stuart (1979) Structure Homology program allowing the superposition of structurally related molecules.

DSSP[2]: Kabsch and Sander (1983) Protein secondary structure prediction via pattern recognition of hydrogen-bonded and geometric features.

Graphical programs

FRODO[5]: Jones (1985). Coordinate and electron density map manipulation.

O[4]: Jones *et al.* (1991). Improved methods for building protein models in electron density maps and the location of errors in these models.

Implementation

1: MicroVAX 4000

2: MicroVAX 3300

3: Silicon Graphics Indigo

4: Evans and Sutherland ESV-10

5: PS390

Chapter 3 : Purification of *Thermoplasma acidophilum* citrate synthase expressed in *Escherichia coli*

3.1 INTRODUCTION

Expression of a thermostable protein in a mesophilic host allows thermal denaturation to be used as a basis for purification of the expressed protein. A purification protocol developed by Sutherland *et al.* (1991) was used for the basic strategy of purifying *Tp.acidophilum* citrate synthase.

3.2 RESULTS

3.2.1 Thermal denaturation of host *E.coli* proteins

A cell extract was made from a 1l culture of *E.coli* TG-1 cells expressing the pCSEH19 clone and was incubated at $65°C$ for 15min. A 5-fold purification of *Tp.acidophilum* citrate synthase was seen after incubation, with a slight loss in enzymic activity. The citrate synthase was then purified to homogeneity using 2 alternative methods.

3.2.2 Purification of *Tp.acidophilum* citrate synthase using chromatofocussing

The heat-treated extract was subjected to gel filtration on a FPLC Superdex 200 column(Pharmacia-LKB), pre-equilibrated with 25mM Triethylamine (pH11.0). The fractions containing enzymic activity were pooled and spun through a Microsep centrifugal concentrator (30K cutoff) until a small sample volume (eg. 6ml) had been reached.

The concentrated extract was then loaded onto a PBE118 chromatofocussing column on a Pharmacia-LKB FPLC system and eluted with Pharmalyte (pH8.0) on a descending pH gradient from pH10.5-8.0. *Tp.acidophilum* citrate synthase activity eluted from the column over a pH range of pH 8.7-

8.3. The active fractions were pooled and concentrated (as before) to a volume of 2ml. This sample was then gel filtered again using the FPLC Superdex 200, this time pre-equilibrated with 20mM Tris-HCl (pH8.0), the buffer used for crystallization trials (see Chapter 5). The citrate synthase eluted in 10ml which was then concentrated (as before) to a final protein concentration of 20mg.ml^{-1}. The sample gave a single band on SDS-polyacrylamide gel electrophoresis (Fig. 3.1) with a M_r 42 000 (+/−2 000). A single band was also seen on an IEF gel (Fig. 3.2), which corresponds to a protein of P_I=8.5 (+/−0.5). Table 3.1 is a summary of the results from a standard purification.

3.2.3 Purification of *Tp.acidophilum* citrate synthase using Dye-ligand Chromatography

The heat-treated extract was loaded onto a 20ml column packed with Matrex Red GelA using 20mM Tris-HCl, 2mM EDTA buffer (pH8.0). The bound citrate synthase was specifically eluted with a combination of substrate and product (5mM OAA and 1mM CoA) in the loading buffer. Six 1ml fractions contained the majority of *Tp.acidophilum* citrate synthase activity and these were pooled and concentrated (as before) to a 2ml volume. This sample was loaded onto a FPLC Superdex 200 gel filtration column pre-equilibrated with 20mM Tris-HCl (pH8.0). The active fractions were concentrated (as before) to a final protein concentration of 20mg.ml^{-1}. The sample gave a major band on an SDS-polyacrylamide gel electrophoresis (Fig. 3.3) with M_r 42 000 (+/− 2 000) and a very minor contaminant of higher molecular weight. A doublet was seen on an IEF gel (not shown), which corresponds to a P_I of around 8.5 (+/− 0.5). Table 3.2 is a summary of the results from a standard purification.

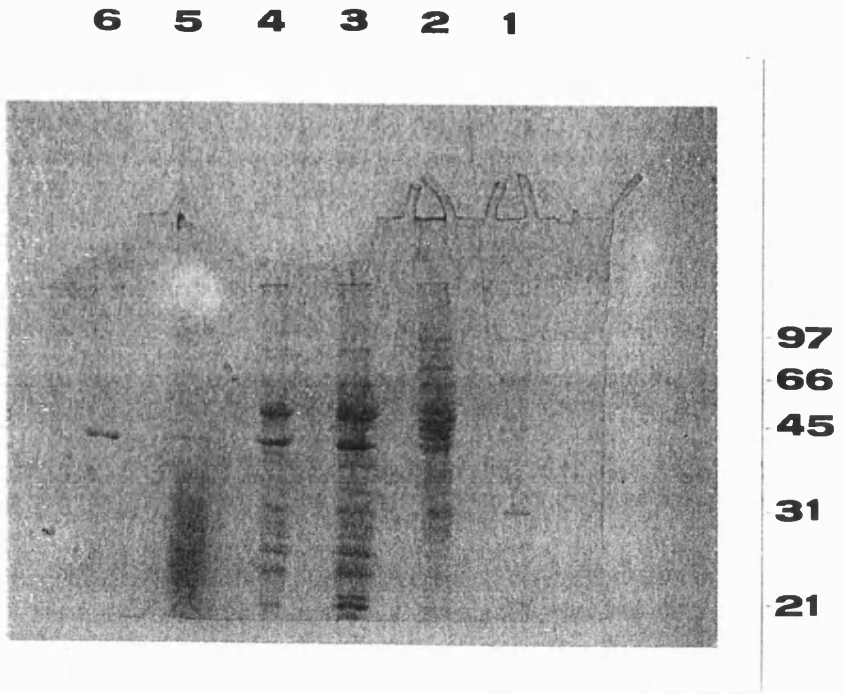


Figure 3.1: SDS-polyacrylamide gel of the stages employed in the purification of *Tp.acidophilum* citrate synthase from *E.coli* TG1 containing pCSEH19. Lane1: Standard molecular weight markers. Lane2: Crude extract. Lane3: Heat-treated extract. Lane4: After gel filtration on Superdex 200 using FPLC (Pharmacia-LKB). Lane5: After chromatofocussing using PBE118 (Pharmacia-LKB). Lane6: After gel filtration on Superdex 200 using FPLC (Pharmacia-LKB). Standard molecular weight markers: Rabbit Muscle Phosphorylase b 97 400, Bovine Serum Albumin 66 200, Hen Egg White Ovalbumin 45 000, Bovine Carbonic Anhydrase 31 000 and Soy Bean Trypsin Inhibitor 21 500.

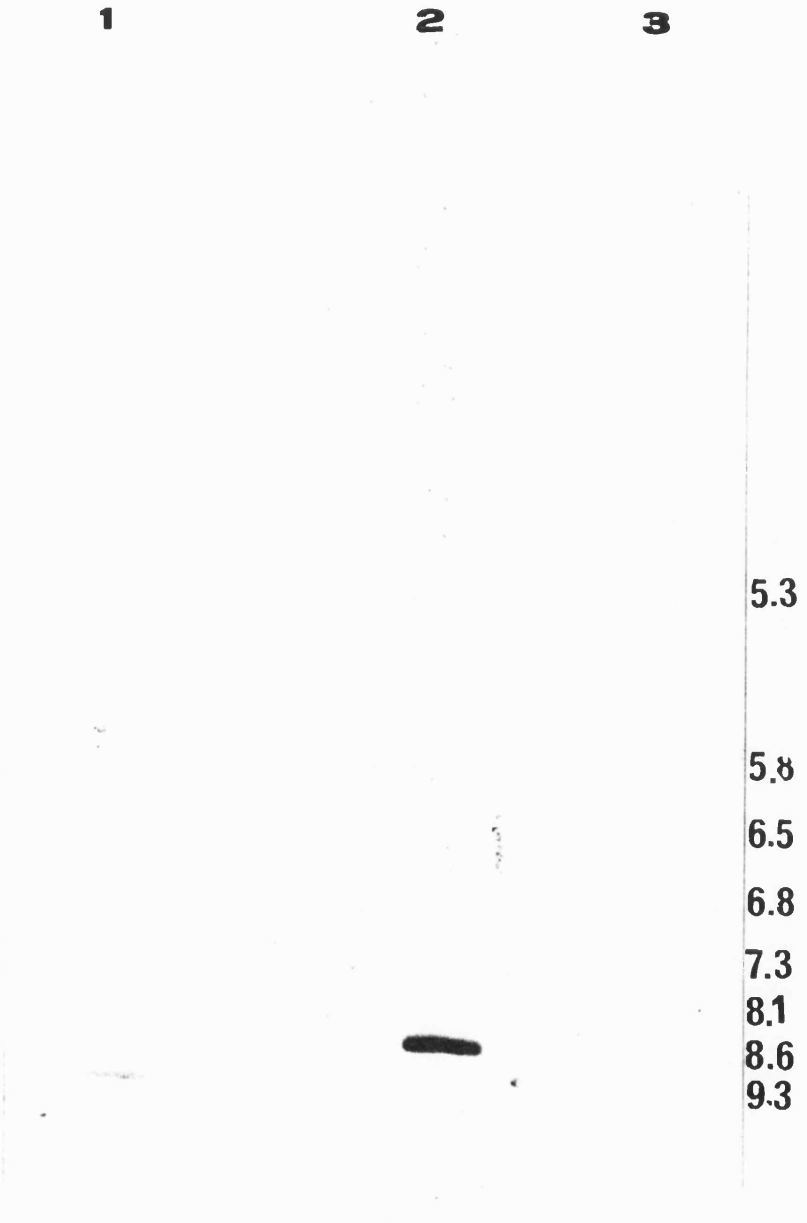


Figure 3.2: Isoelectric focussing (IEF) Phastgel (pH 3-9). Lane1: Isoelectric focussing markers. Lane2: Purified *Tp.acidophilum* citrate synthase, via chromatofocussing. Lane3: Isoelectric focussing markers. Isoelectric focussing markers: β-lactoglobulin 5.3, Bovine Carbonic Anhydrase B 5.8, Human Carbonic Anhydrase B 6.5, Horse Myoglobin 6.8, Horse Myoglobin 7.3, Lentil Lectin 8.1, Lentil Lectin 8.6 and Trypsinogen 9.3.

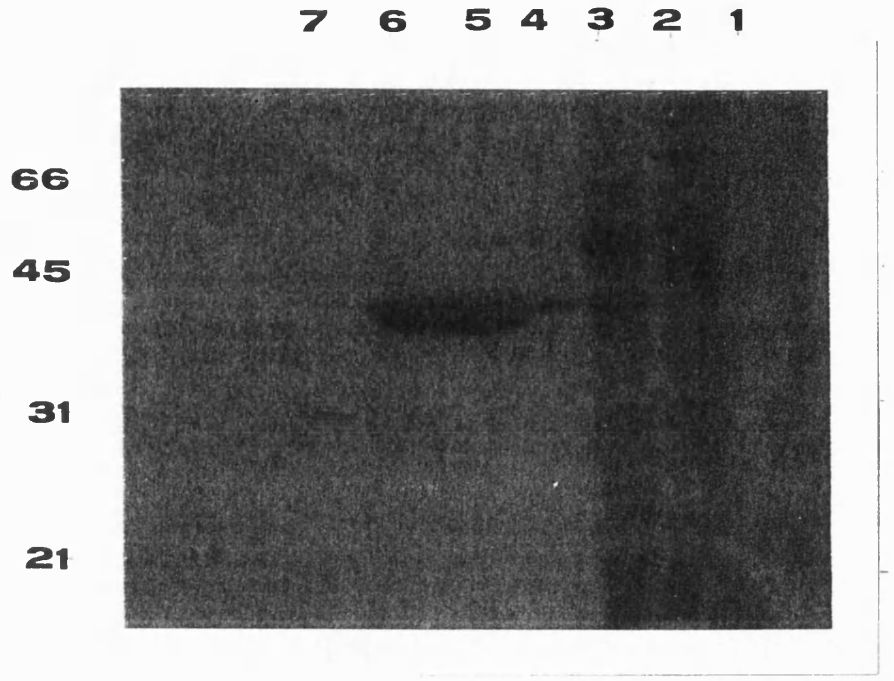


Figure 3.3: SDS-polyacrylamide gel of the stages employed in the purification of *Tp.acidophilum* citrate synthase from *E.coli* TG1 containing pCSEH19. Lane1: Standard molecular weight markers. Lane2: Crude extract. Lane3: Heat-treated extract. Lane4: After elution from Matrex Red GelA column with 5mM oxaloacetate and 0.2mM CoA. Lane5: After gel filtration on Superdex 200 using FPLC (Pharmacia-LKB). Lane6: After gel filtration on Superdex 200 using FPLC (Pharmacia-LKB). Lane7: Standard molecular weight markers. Standard molecular weight markers: Bovine Serum Albumin 66 200, Hen Egg White Ovalbumin 45 000, Bovine Carbonic Anhydrase 31 000 and Soy Bean Trypsin Inhibitor 21 500

step	total protein (mg)	total activity (units)	specific activity (units/mg)	purification (fold)	yield (%)
Cell extract	560	1183	2.1	–	100
Heat treatment	80	840	10.5	5	71
Superdex 1	50	560	11.2	5.3	47
PBE118	14	330	23.6	11.2	28
Superdex 2	7	285	40.7	19.4	24

Table 3.1: Purification of *Tp.acidophilum* citrate synthase from *E.coli* TG1 containing pCSEH19 using chromatofocussing

step	total protein (mg)	total activity (units)	specific activity (units/mg)	purification (fold)	yield (%)
Cell extract	400	762	1.9	–	100
Heat treatment	70	620	8.9	4.7	86
Matrex Red GelA	8	310	38.8	20.4	40
Superdex	6	240	40.7	21.4	31

Table 3.2: Purification of *Tp.acidophilum* citrate synthase from *E.coli* TG1 containing pCSEH19 using Dye ligand chromatography

3.3 DISCUSSION

The expression of a thermostable protein in a mesophilic host has allowed thermal denaturation of host proteins to be an important and rapid step in the purification step of *Tp.acidophilum* citrate synthase. Although *E.coli* is a mesophile, some of its proteins show thermotolerance leading to only partial purification of the expressed thermostable protein. The overexpression also allows large scale production of the protein in sufficient quantities for crystallization trials. Two purification protocols following the heat denaturation step have been devised, each using a different type of chromatographic procedure.

3.3.1 Chromatofocussing

The gel filtration of the heat-treated *E.coli* cell extracts does not significantly purify the protein but it allows the simple transfer of the sample into the correct buffer for the chromatofocussing step, which is very dependent on having the sample at exactly the right pH value.

Chromatofocussing was an efficient and rapid method for purifying *Tp.acidophilum* citrate synthase. The citrate synthase was one of the first proteins to be eluted from the column at pH 8.7-8.3, indicating the high isoelectric point of the protein. The problem with chromatofocussing is the use of Pharmalyte as the eluent. Although the protein is pure after this stage, Pharmalyte is also present in the sample and appears as a large smear on SDS-PAGE gels (Fig. 3.1). As the aim of this purification is to gain highly pure protein for crystallization trials, the presence of Pharmalyte in the sample would be deleterious. The Pharmalyte was removed by gel filtration. This final step also allows the transfer of the protein into a suitable buffer for crystallization trials, ie. 20mM Tris-HCl (pH8.0).

The purified citrate synthase had a specific activity of 41 units/mg

protein compared to 41 units/mg protein when the enzyme was purified from *Tp.acidophilum* (Smith *et al.*, 1987), indicating similar levels of specific activity and purity from the expressed and native enzyme.

This purification procedure yields 6-7mg *Tp.acidophilum* citrate synthase from a 1l culture of *E.coli*, an amount suitable for wide-spread crystallization trials. The protein is a single band on both SDS-PAGE and IEF gels, a condition which is very advantageous for the sometimes problematic protein crystallization.

3.3.2 Dye ligand chromatography

Matrex Red GelA can be used as an affinity purification step for proteins containing an adenine nucleotide attached to a phosphorylated ribose binding site. Citrate synthase has a binding site for CoA, which consists partly of an adenine ring attached to a phosphorylated ribose, and so should bind to the matrix. Other citrate synthases (eg from *E.coli* (Weitzman and Ridley, 1983)) have been reported to bind to Red Gel , but they have been eluted non-specifically with KCl. In the present case, a combination of substrate (oxaloacetate) and product (CoA) was used to try and elute the bound citrate synthase specifically. After loading of the heat-treated extract from a 500ml culture of *E.coli* onto the dye column, no citrate synthase activity was seen in the wash, indicating that the enzyme had bound to the column. Under the elution conditions used, 50% of the loaded enzymic activity was recovered. This sample now contains oxaloacetate and CoA, which are not wanted for the crystallization of the native *Tp.acidophilum* citrate synthase. These were removed by gel filtration.

The specific activity for *Tp.acidophilum* citrate synthase found was 41 units/mg protein after the gel filtration, indicating that the two purification procedures yield protein of similar purity. The contamination seen on a

SDS-PAGE gel is very minor and can only be visualised with Coomassie Blue when the gel was overloaded, indicating that elution using oxaloacetate and CoA is highly specific. The doublet seen on an IEF gel may be due to the heat treatment causing minor deamidation and thus leading to a slight charge heterogeneity. Although the level of purity was not ideal it was deemed pure enough for crystallization trials.

This purification procedure yields 6-7mg *Tp.acidophilum* citrate synthase from a 1l culture, an amount suitable for wide spread crystallization trials.

3.3.3 The better of the purification procedures?

Both of the above purification procedures yield roughly the same amount of *Tp.acidophilum* citrate synthase from a 1l culture, with comparable specific activities. However, there are minor differences. Firstly, the Dye ligand protocol is shorter than the chromatofocussing protocol, having one less column run to undertake. Secondly, the level of purity gained. With chromatofocussing a single band is seen on SDS-PAGE compared to the slight contaminant present using Matrex Red GelA. Also with chromatofocussing a single band is seen on an IEF gel compared to the doublet seen using Matrex Red GelA. This charge homogeneity with the former method is probably due to the fact that chromatofocussing is separating proteins due to their charge. Therefore, since protein homogeneity is a key factor in protein crystallography, the chromatofocussing method is preferable, although crystals have been grown from protein purified using the Dye ligand method (see Chapter 5).

Chapter 4 : Characterisation of *Thermoplasma acidophilum* citrate synthase.

4.1 INTRODUCTION

The study into the basis of the inherent stability of some proteins has concentrated recently on the determination of the protein's crystal structure and on mutagenesis experiments, usually based on the crystal structure, to alter the protein's stability (see Chapter 1). A prerequisite for site-specific mutagenesis experiments to investigate, in this case, thermal stability is a reproducible way of characterising the mutant in relation to the wild type, the wild types being *Tp.acidophilum* citrate synthase and pig heart citrate synthase.

Assays based on the loss of enzymic activity and on changes in the circular dichroism spectrum at known temperatures have been developed that will quantitatively measure possible differences between the wild type and specific mutants that have been generated with the aim of producing a protein with altered thermal stability. Although no site-directed mutagenesis experiments have been performed the work presented below and throughout this thesis will lay the foundations for such experiments in the future.

4.1.1 The Arrhenius plot of thermal inactivation

To investigate the thermal stability of *Tp.acidophilum* citrate synthase, the rates of inactivation at known temperatures were determined. To quantify the rates of a reaction in relation to temperature the Arrhenius equation can be used.

$$\frac{d\ ln\ k}{d\ T} = \frac{E_a}{RT^2}$$

where k is the rate constant for the reaction, E_a is the activation energy,

R is the gas constant and T is the temperature in degrees Kelvin. E_a can be assumed to be independent of temperature; thus, integrating the above equation gives:

$$k = A\ e^{-\frac{E_a}{RT}}$$

where A is known as the pre-exponential factor.

A plot of ln k vs. 1/T, known as the Arrhenius plot, gives a straight line, the gradient of which is $-E_a$/R. For a thermal inactivation reaction, E_a can be seen as the energy of activation of the inactivation process.

4.1.2 Circular dichroism spectroscopy

To characterise further the stability of wild type *Tp.acidophilum* citrate synthase circular dichroism (CD) studies were performed on the purified enzyme. The CD spectra were performed using a JASCO J-600 spectropolarimeter by Dr N.C. Price at the University of Stirling, in collaboration with Dr. D. W. Hough.

CD is a form of absorption spectroscopy that uses circularly-polarised light instead of unpolarised light. An asymmetric molecule, such as a protein, will absorb right circularly polarised light differently from left circularly polarised light. This leads to the phenomenom of CD which is the difference in extinction coefficients between right and left circularly polarised light. The CD spectrum of a protein is composed of the linear combination of spectra for the component secondary structures. α-Helices, β-sheets and β-turns have characteristic CD spectra. Thus, since CD is very sensitive to secondary structural components of proteins, the CD spectrum of a protein will give information on the amount of different types of secondary structure that makes up its conformation.

Results are presented below showing how CD spectroscopy has been used to investigate both the thermal unfolding and Gdn-HCl induced un-

folding of *Tp.acidophilum* citrate synthase.

4.2 RESULTS

4.2.1 Thermal inactivation of *Tp.acidophilum* citrate synthase

The inactivation experiment was carried out using the method detailed in section 2.2.10 over a time course of 10min at known incubation temperatures. The value for 100% activity was gained by extrapolating back to time zero the time courses where no inactivation was seen ($55°C$ - $70°C$). The inactivation plot at $81°C$ is shown in Fig.4.1. The inactivation profile over a temperature range can be seen in Fig.4.2, where the % activity remaining after 10min is plotted against temperature.

The Arrhenius plot for the thermal inactivation of *Tp.acidophilum* citrate synthase is shown in Fig.4.3. Two characteristics of the wild type can be gained from this plot. Firstly, for *Tp.acidophilum* citrate synthase E_a = 364 KJmol^{-1} +/- 72 KJmol^{-1} and secondly, the position of the line in relation to the x-axis, ie the temperature axis, which relates to the temperature over which inactivation is seen.

All gradients for the inactivation plots and Arrhenius plots were calculated using linear regression using Quattro Pro, with each data point in the graphs being the mean value from triplicate experiments. Error values for the Arrhenius plot gradient were obtained from the linear regression calculation.

In order to highlight the relative thermal stability of *Tp.acidophilum* citrate synthase compared to the enzyme from other organisms, thermal characterisation of pig heart and *Sulfolobus solfataricus* citrate synthases were performed.

The same method of thermal characterisation as for *Tp.acidophilum*

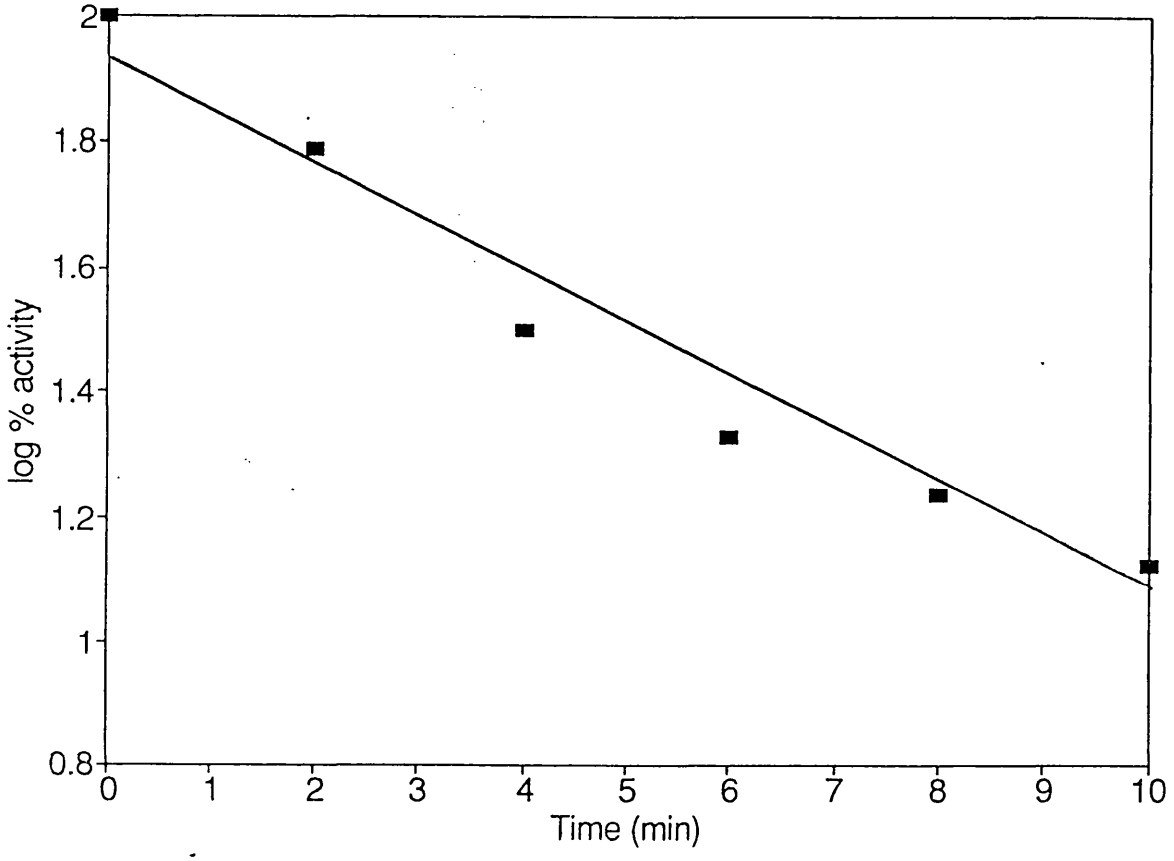


Figure 4.1: Time course for the thermal inactivation of *Tp.acidophilum* citrate synthase at 81°*C*. Assay buffer was equilibrated at the required temperature, the enzyme sample added and an aliquot assayed for enzymic activity after a known time. The fraction of enzymic activity present is plotted as log% of initial activity, which was gained by extrapolation back to zero on the time courses where no inactivation was seen.

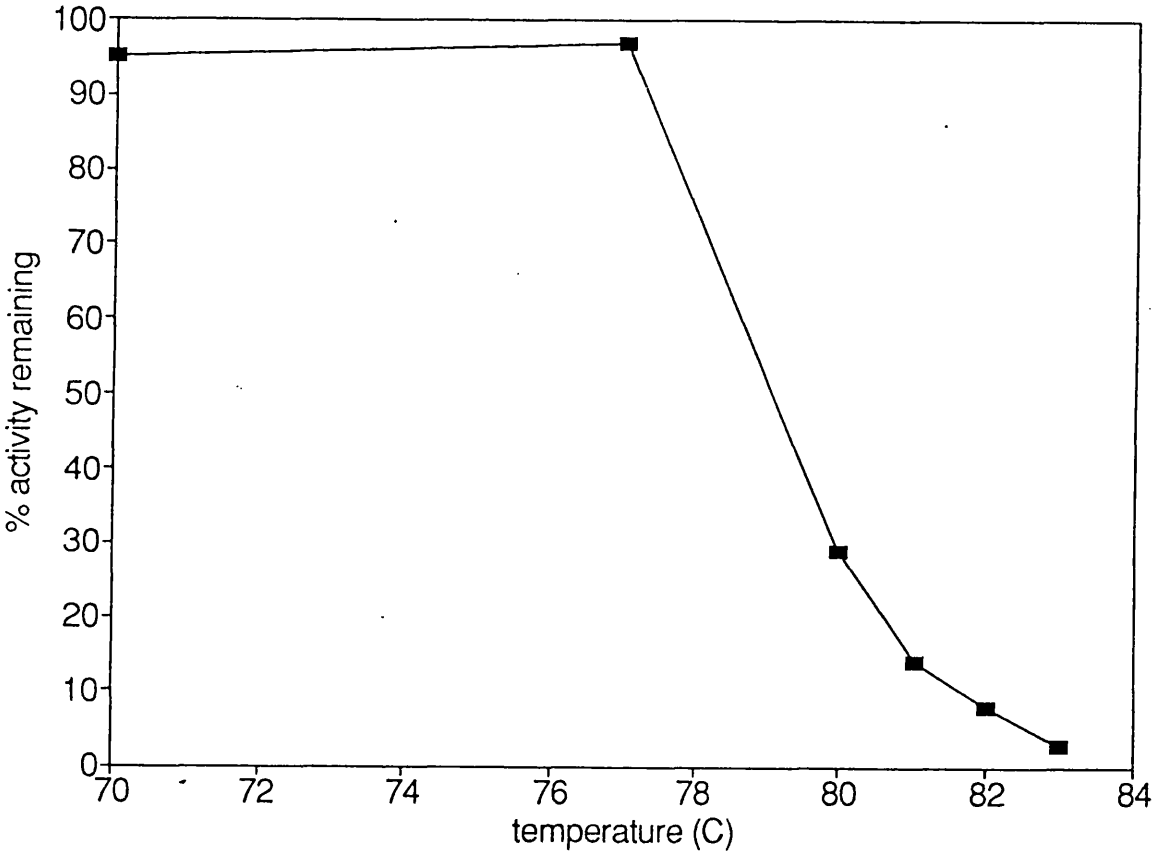


Figure 4.2: The thermal inactivation profile for *Tp.acidophilum* citrate synthase. The % enzymic activity remaining of the initial activity after incubation for 10min at a specific temperature.

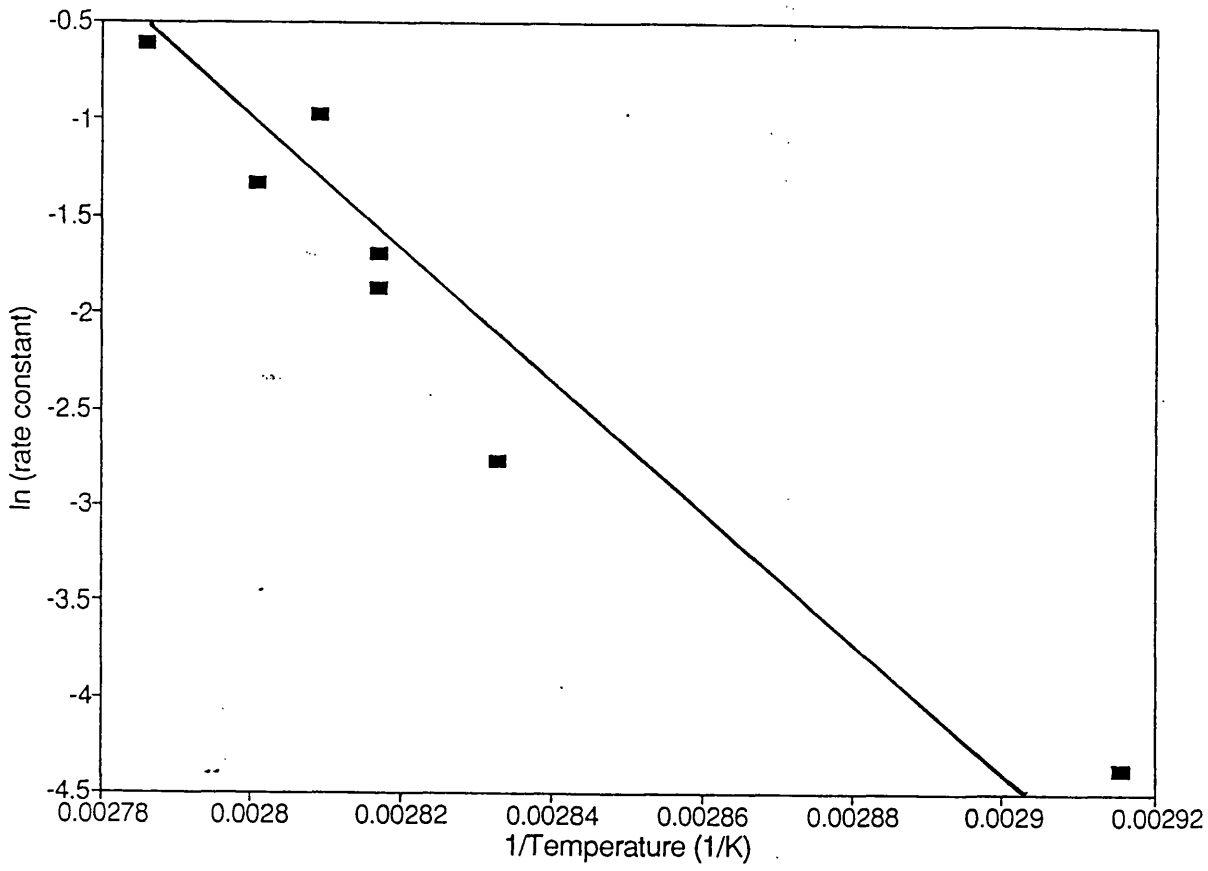


Figure 4.3: The Arrhenius plot for the thermal inactivation of *Tp.acidophilum* citrate synthase. The ln of the rates of thermal inactivation at a specific temperature are plotted against the reciprocal of those temperatures in degrees Kelvin. The gradient is equivalent to the energy of activation of the inactivation process divided by the gas constant, R.

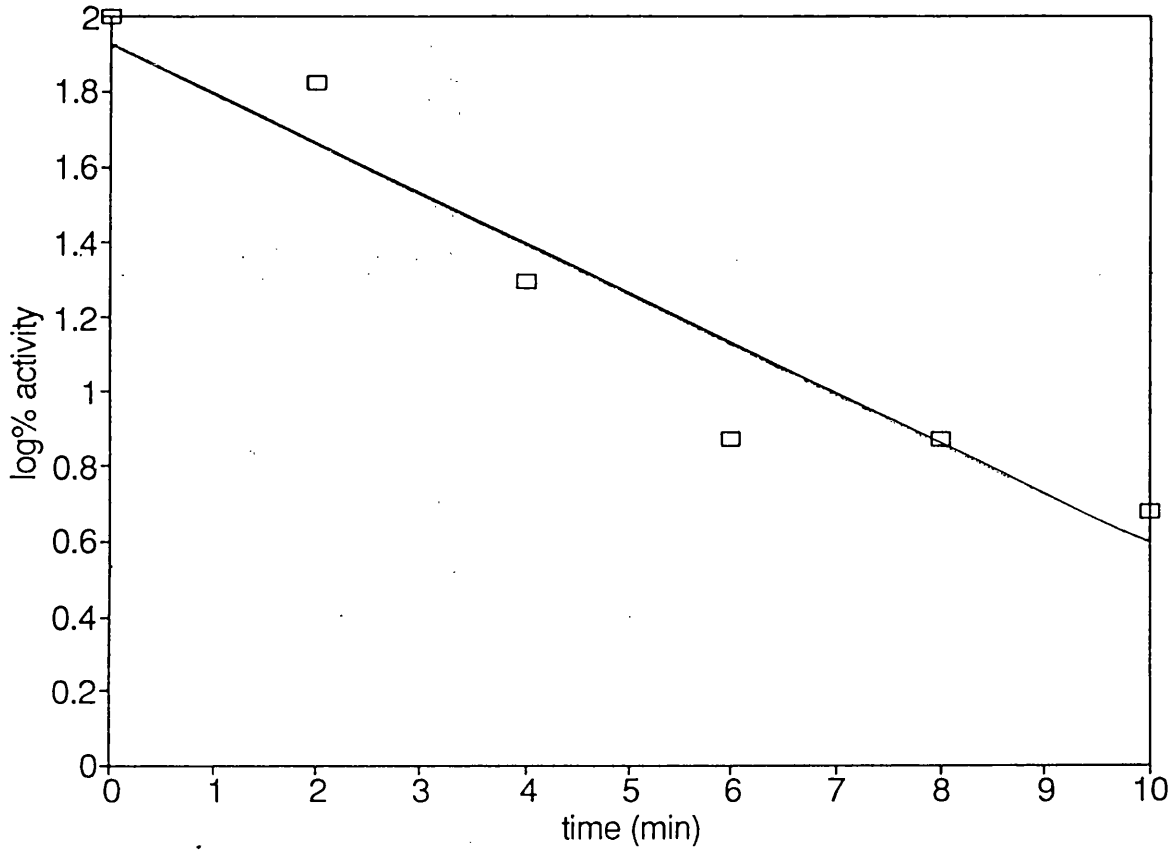


Figure 4.4: Time course for the thermal inactivation of pig heart citrate synthase at 45°C. Assay buffer was equilibrated at the required temperature, the enzyme sample added and an aliquot assayed for enzymic activity after a known time. The fraction of enzymic activity present is plotted as log% of initial activity, which was gained by extrapolation back to zero on the time courses where no inactivation was seen.

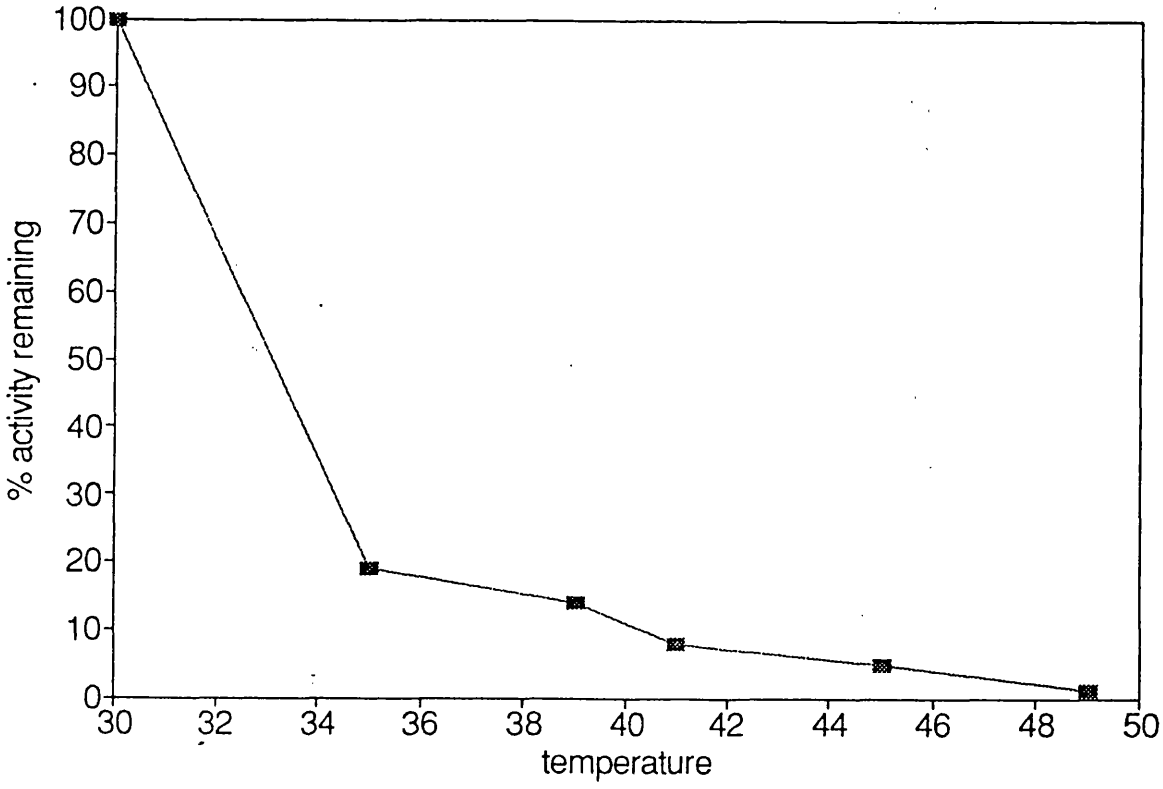


Figure 4.5: The thermal inactivation profile for pig heart citrate synthase. The % enzymic activity remaining of the initial activity after incubation for 10min at a specific temperature.

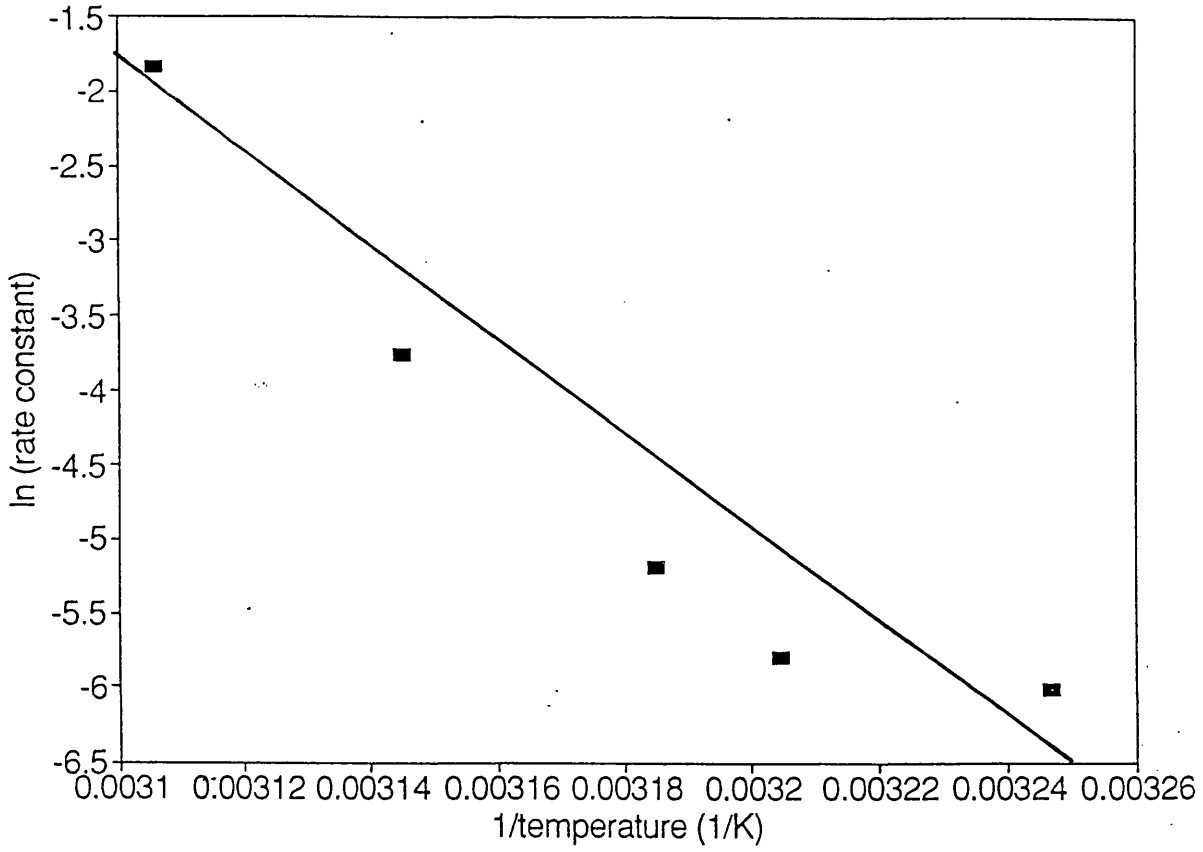


Figure 4.6: The Arrhenius plot for the thermal inactivation of pig heart citrate synthase. The ln of the rates of thermal inactivation at a specific temperature are plotted against the reciprocal of those temperatures in degrees Kelvin. The gradient is equivalent to the energy of activation of the inactivation process divided by the gas constant, R.

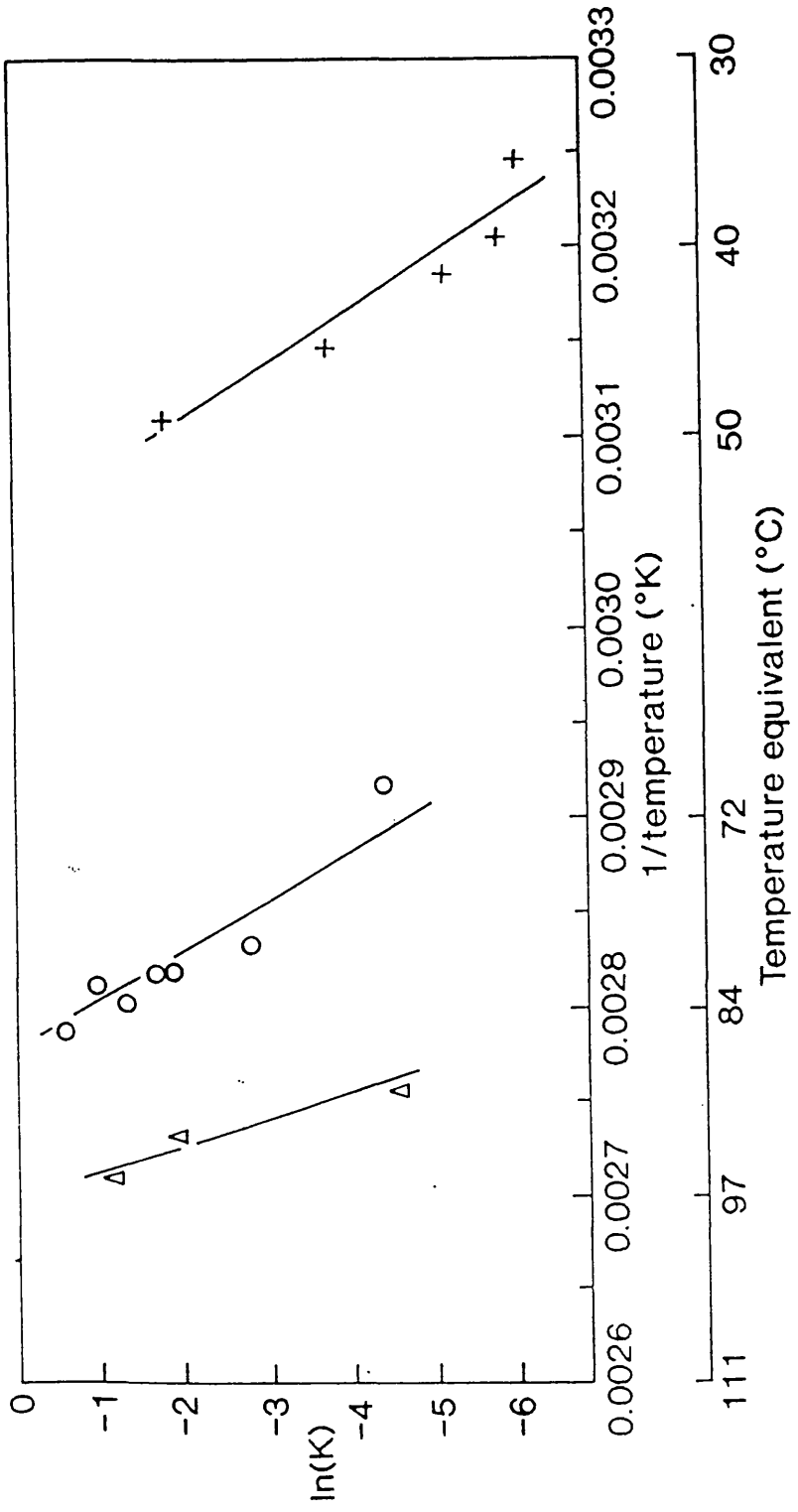


Figure 4.7: A comparative Arrhenius plot (ln rate constant vs. 1/temperature) for pig heart, *Tp.acidophilum* and *Sulfolobus solfataricus* citrate synthases, displaying the differing thermal stability of the citrate synthases from these organisms.

Cross: Pig heart citrate synthase

Circle: *Tp.acidophilum* citrate synthase

Triangle: *S.sulfataricus* citrate synthase

citrate synthase was used here. Fig.4.4. shows the time dependent inactivation and Fig.4.5. shows the temperature-dependent inactivation profile of pig heart citrate synthase. Fig.4.6. shows the Arrhenius plot for the thermal inactivation of pig heart citrate synthase. E_a for pig heart citrate synthase = 331KJmol^{-1} +/- 22 KJmol^{-1}. The comparative Arrhenius plots for *Tp.acidophilum* and pig heart citrate synthase are shown in Fig.4.7. Thermal inactivation data in the form of an Arrhenius plot for citrate synthase from *Sulfolobus solfataricus*, a thermophilic archaeon growing optimally at 85°C is also presented.

4.2.2 Circular dichroism spectroscopy studies of *Tp.acidophilum* citrate synthase

CD spectra were recorded in Tris buffer (pH8.0) from 190nm-250nm. The CD spectrum for *Tp.acidophilum* citrate synthase was taken at 20°C and at 55°C - see Fig.4.8 , over the wavelength 190nm–250nm. The spectra are very similar indicating no large scale structural differences at the two temperatures.

The estimation of the percentage of secondary structure found in *Tp.acidophilum* citrate synthase from the CD spectra was calculated by two methods. The first method is that of Provencher and Glöckner (1991) using the program CONTIN, which uses a reference set of well-characterised proteins for the prediction ; the second method is that of Siegel *et al* (1980) which uses a narrower range of wavelengths to predict only the α-helical content.

Prediction type	% α-helix	% β-sheet	% other
CONTIN	54	41	5
Siegel	53	–	–

To test how good the prediction methods are, the CD spectrum for pig heart citrate synthase was taken and the predicted % secondary structure

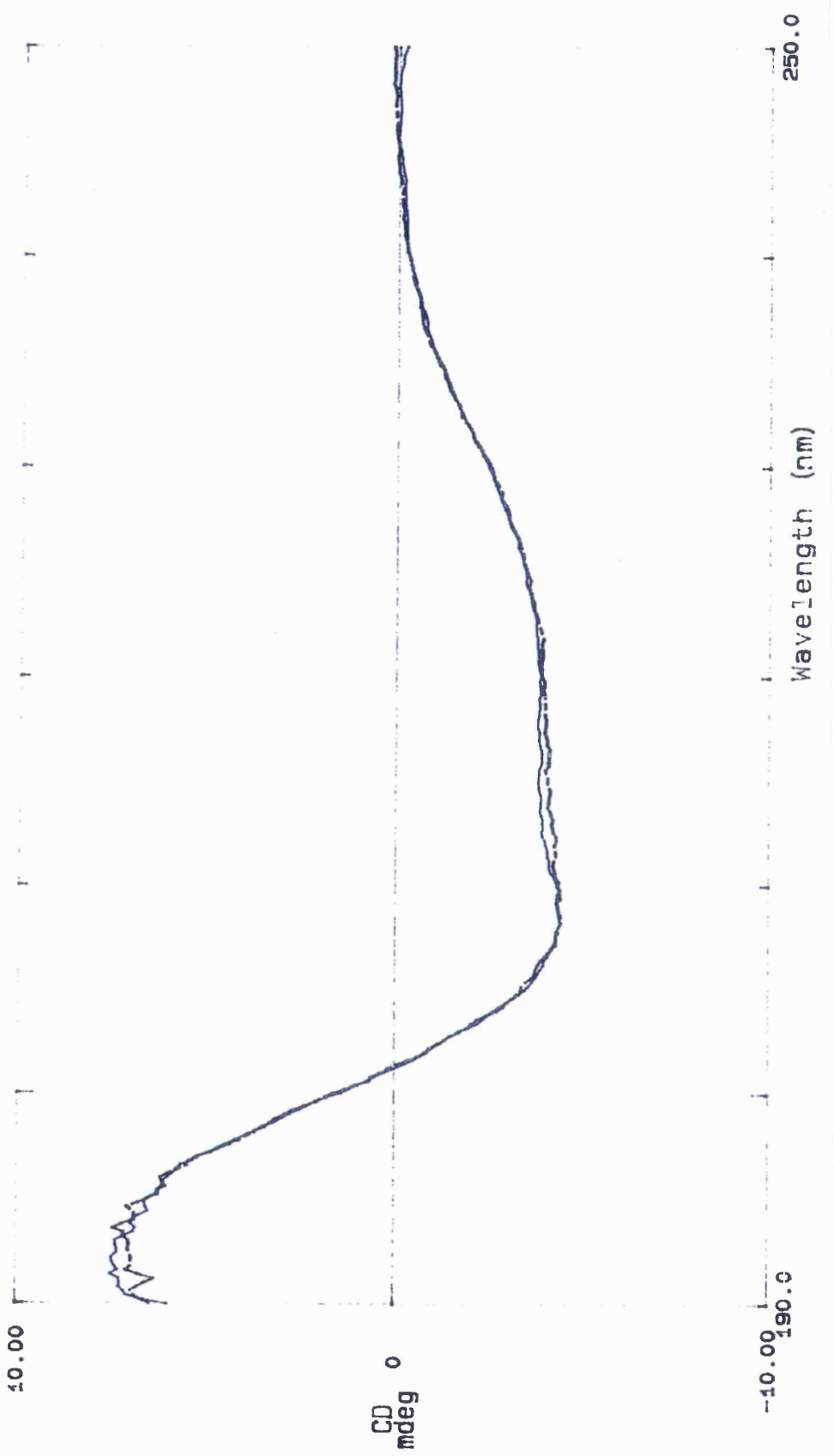


Figure 4.8: Far u.v. circular dichroism spectra of *Tp.acidophilum* citrate synthase recorded at 20°C (blue) and 55°C (black) in 10mM Tris-HCl buffer, pH8.0.

values (including a third method by Chang *et al.* (1978)) were compared to the crystal structure values.

Prediction type	% α-helix	% β-sheet	% other
CONTIN	47	49	4
Siegel	53	–	–
Chang	44	33	23

Actual values	% α-helix	% β-sheet	% other
	57	11	32

4.2.3 Thermal denaturation studies of *Tp.acidophilum* citrate synthase

To complement the thermal inactivation studies of *Tp.acidophilum* citrate synthase, CD spectra were taken at higher temperatures to identify at what temperature loss of secondary structure was seen. In order to minimize the time that the sample was cumulatively in the cuvette, the higher temperature data were recorded from 240nm rather than 250nm as before. Spectra were taken at a range of temperatures from $20°C$ – $80°C$ in the cuvette - see Fig.4.9. The spectra for all the temperatures are very similar showing that there is little loss of secondary structure on recording the spectrum at $80°C$. These spectra took 2.5min each to complete, which correlates with the thermal inactivation studies where there was only a small reduction in catalytic activity when the enzyme was incubated at $80°C$ for 2min.

The next set of spectra were recorded after the enzyme had been pre-incubated at a known temperature for a known time - see Fig.4.10. On incubation at $80°C$ for 5min or 20min, *Tp.acidophilum* citrate synthase showed only a small loss of structure, indicated by the shift of the spectra away from the highly negative intensity signal of the 222nm band, indicating loss of α-helical structure. The spectrum taken at $85°C$ with no pre-incubation also showed only a small loss of structure; however after

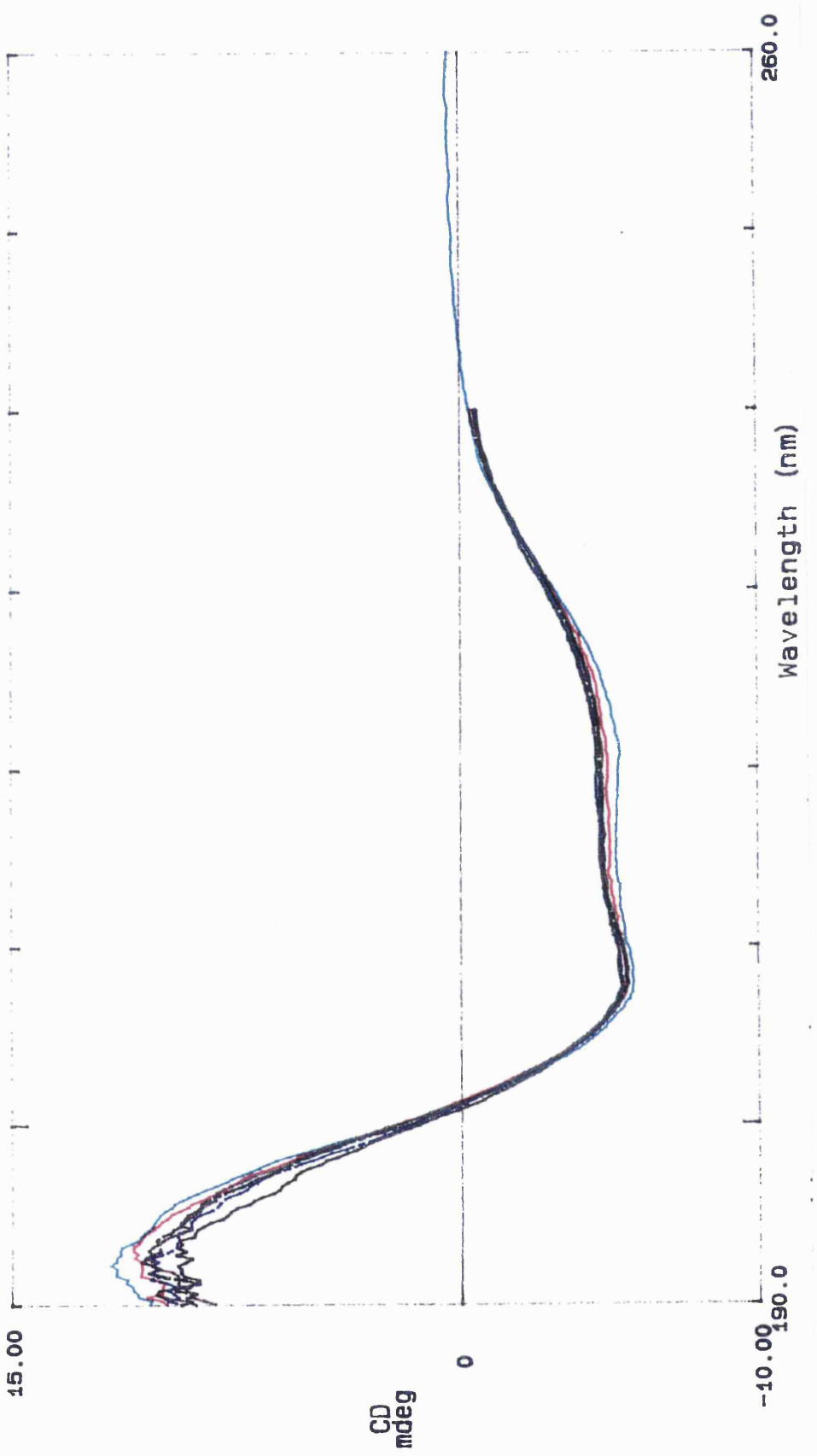


Figure 4.9: Far u.v. circular dichroism spectra of *Tp.acidophilum* citrate synthase recorded at 20°C (green), 55°C (red), 68°C (brown), 73°C (black), 77°C (blue) and 80°C (black) in 10mM Tris-HCl buffer, pH8.0.

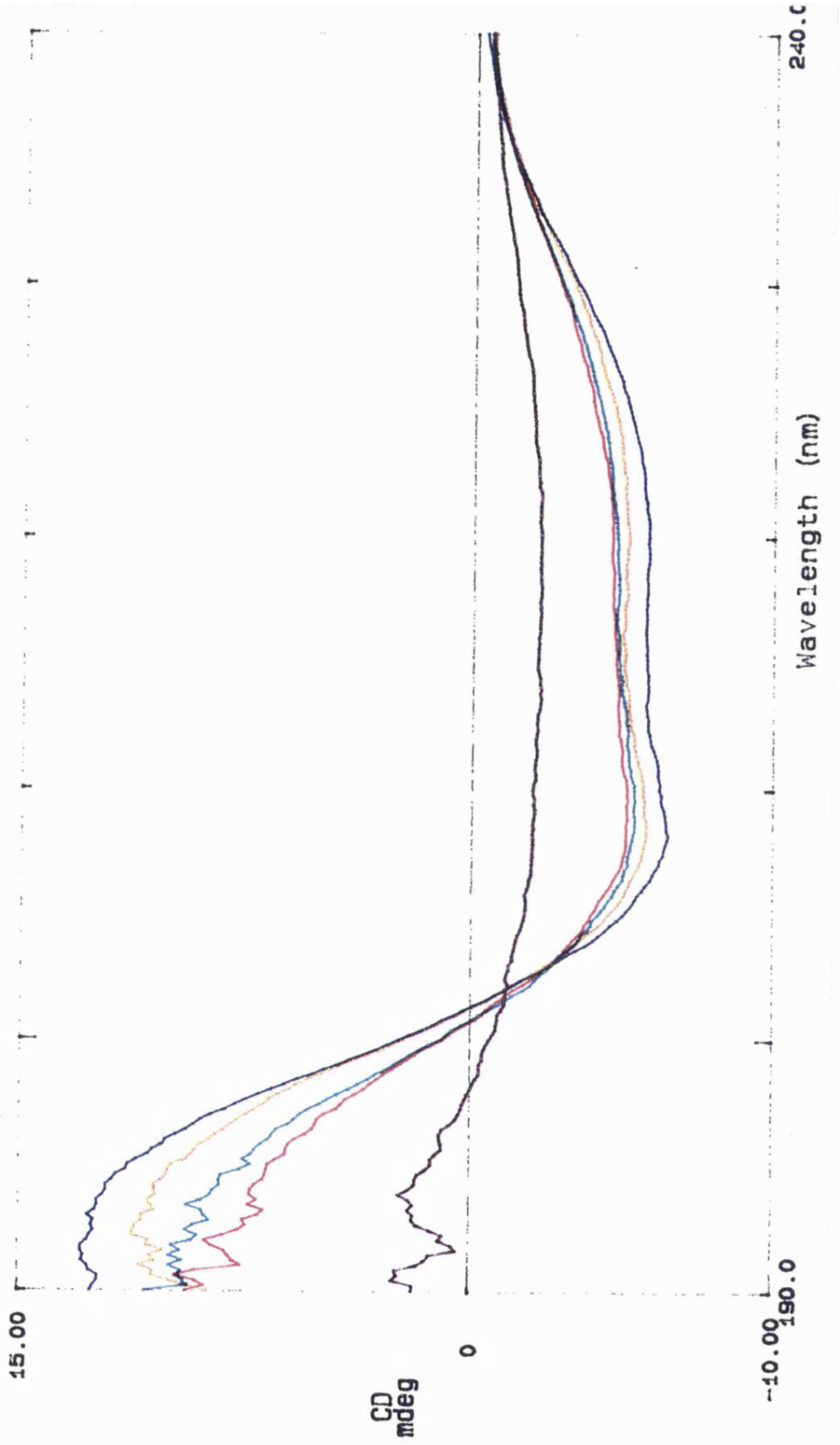


Figure 4.10: Far u.v. circular dichroism spectra of *Tp.acidophilum* citrate synthase recorded at 20°C (blue), 80°C (orange) after 5min pre-incubation at 80°C, 80°C (green) after 20min pre-incubation at 80°C, 85°C (red), 85°C (blue) after 20min pre-incubation at 85°C in 10mM Tris-HCl, pH8.0.

pre-incubation at 85°C for 20min, the spectrum taken at 85°C shows a dramatic change, indicating a drastic loss of secondary structure due to the heat treatment.

To see if this heat denaturation is reversible, a sample of *Tp.acidophilum* citrate synthase was incubated at 85°C for 20min , cooled to 20°C and left overnight at that temperature. A CD spectrum of this sample was then recorded - see Fig.4.11. There is no significant signal present suggesting there is no regain of the structure after heat denaturation, although the spectrum suggests a total loss of material which may be due to the protein having precipitated overnight.

4.2.4 Guanidine-HCl denaturation of *Tp.acidophilum* citrate synthase

To investigate further the stability of the wild type *Tp.acidophilum* citrate synthase, guanidine-HCl unfolding of the enzyme was followed by CD. The spectra were recorded with increasing concentrations of Gdn-HCl in the sample and were carried out at 20°C and at 55°C - see Figs.4.12 and 4.13. The noise below 210nm is due to absorption by Gdn-HCl.

The spectra were analysed by plotting changes in intensity of the signal at 222nm, expressed as % of the total change of α-helical content of the protein, against Gdn-HCl concentration - see Fig.4.14. The data were also compared to the known Gdn-HCl unfolding profile for pig heart citrate synthase at 20°C. Spectra were also taken after the protein was incubated in 6M Gdn-HCl (results not shown). The decrease in signal intensity of these spectra are characteristic of a protein that has totally unfolded under these denaturing conditions.

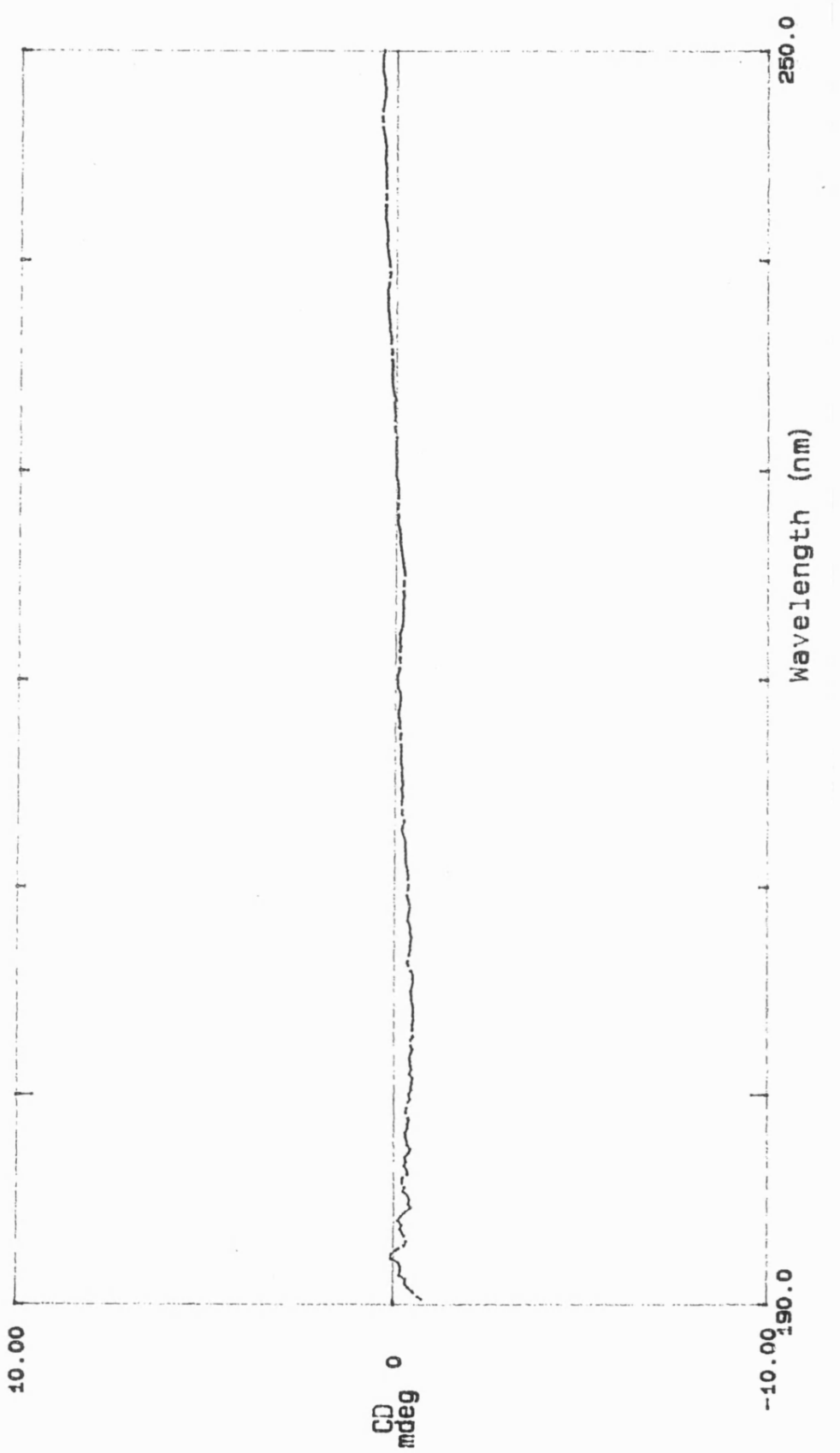


Figure 4.11: Far u.v. circular dichroism spectra of *Tp.acidophilum* citrate synthase recorded at $20°C$, after pre-incubation at $85°C$ for 20min followed by incubation at $20°C$ overnight, in 10mM Tris-HCl buffer, pH8.0

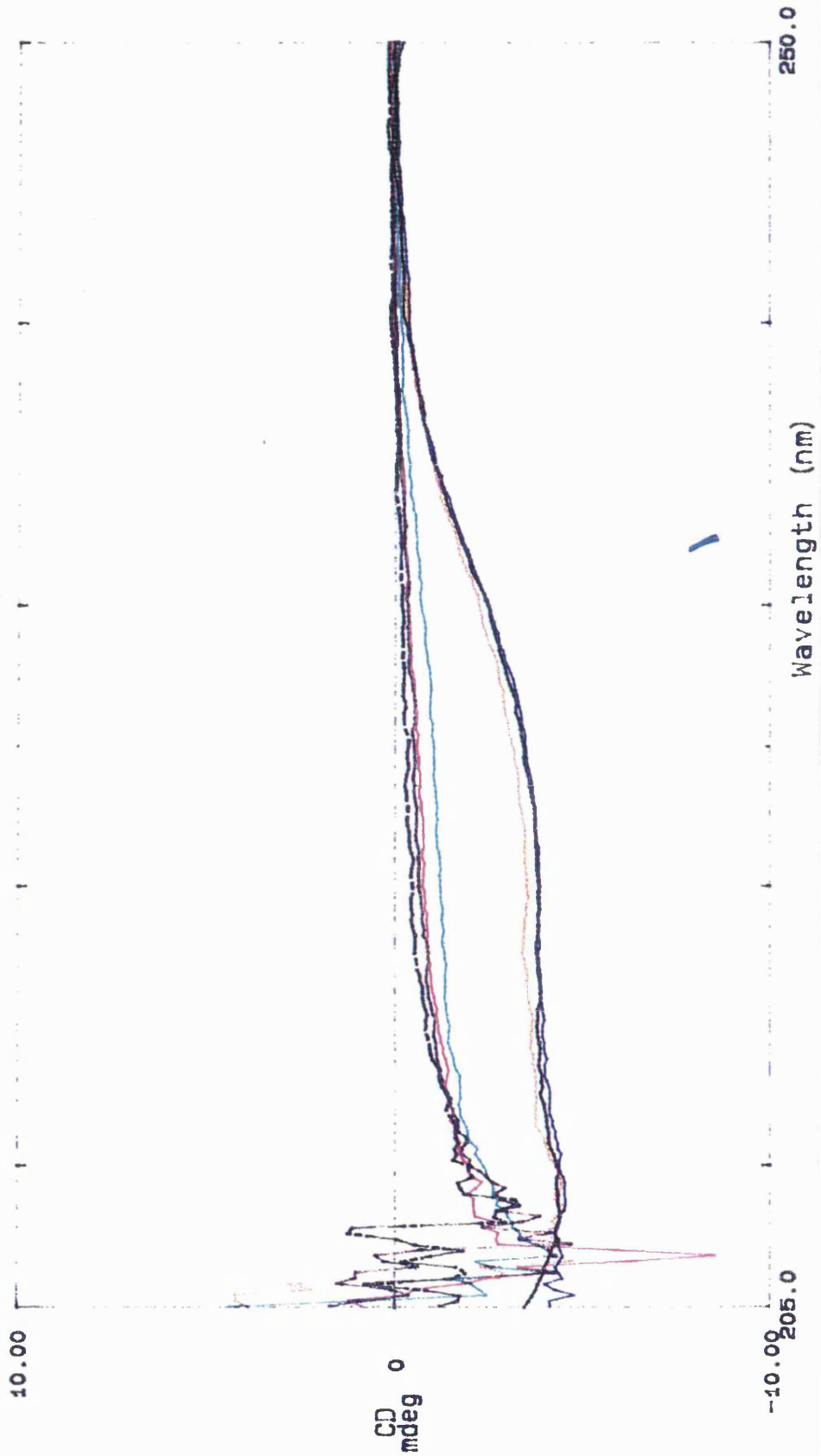


Figure 4.12: Far u.v. circular dichroism spectra of *Tp.acidophilum* citrate synthase recorded at 20°*C* in 0M Gdn-HCl (black), 1M Gdn-HCl (blue), 2M Gdn-HCl (orange), 2.25M Gdn-HCl (green), 2.5M Gdn-HCl (red), 3M Gdn-HCl (brown) and 4M Gdn-HCl (black) in 10mM Tris-HCl buffer, pH8.0.

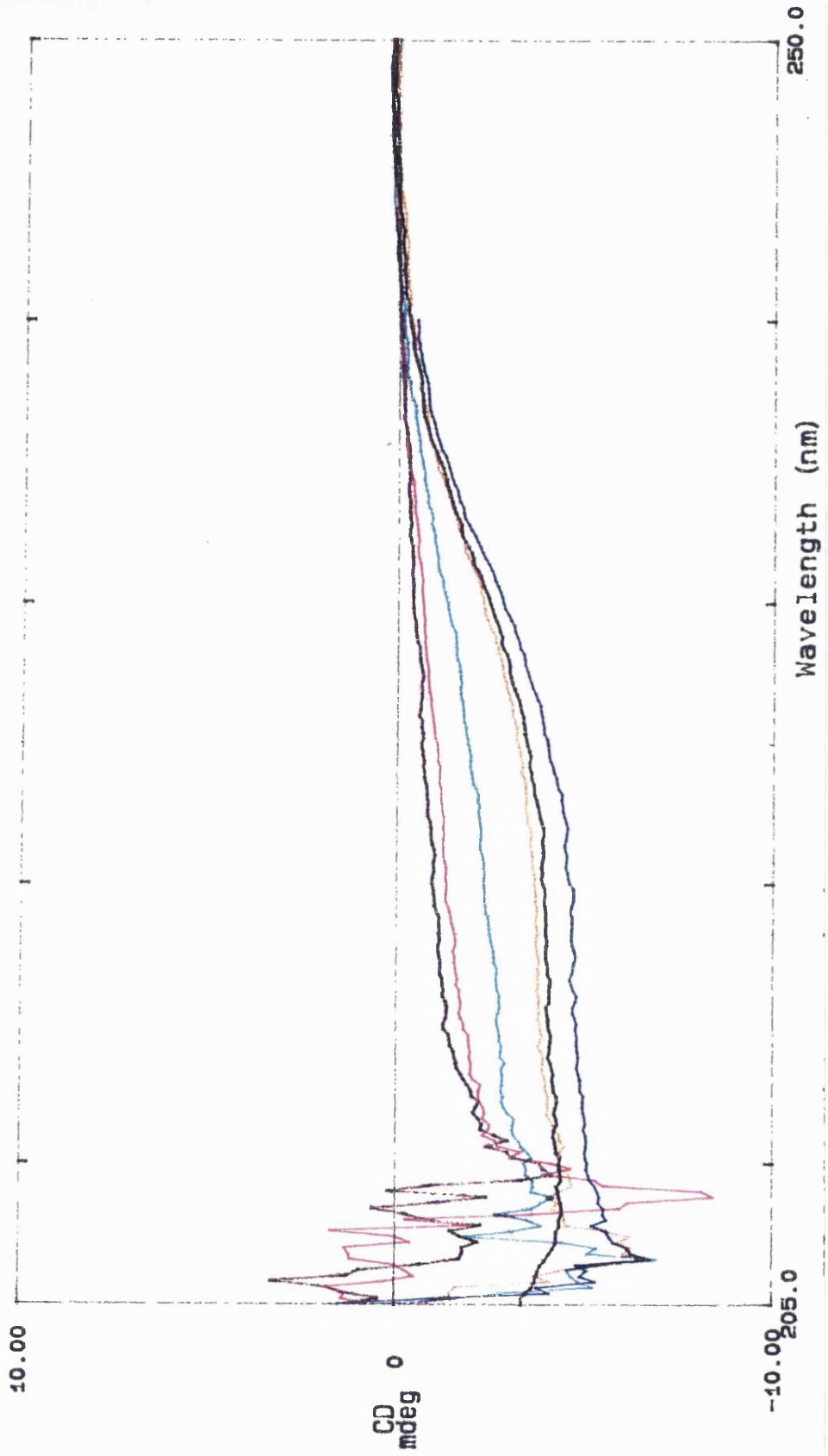


Figure 4.13: Far u.v. circular dichroism spectra of *Tp.acidophilum* citrate synthase recorded at 55°C in 0M Gdn-HCl (black), 1M Gdn-HCl (blue), 1.5M Gdn-HCl (orange), 2.M Gdn-HCl (green), 3M Gdn-HCl (red) and 4M Gdn-HCl (brown) in 10mM Tris-HCl buffer, pH8.0.

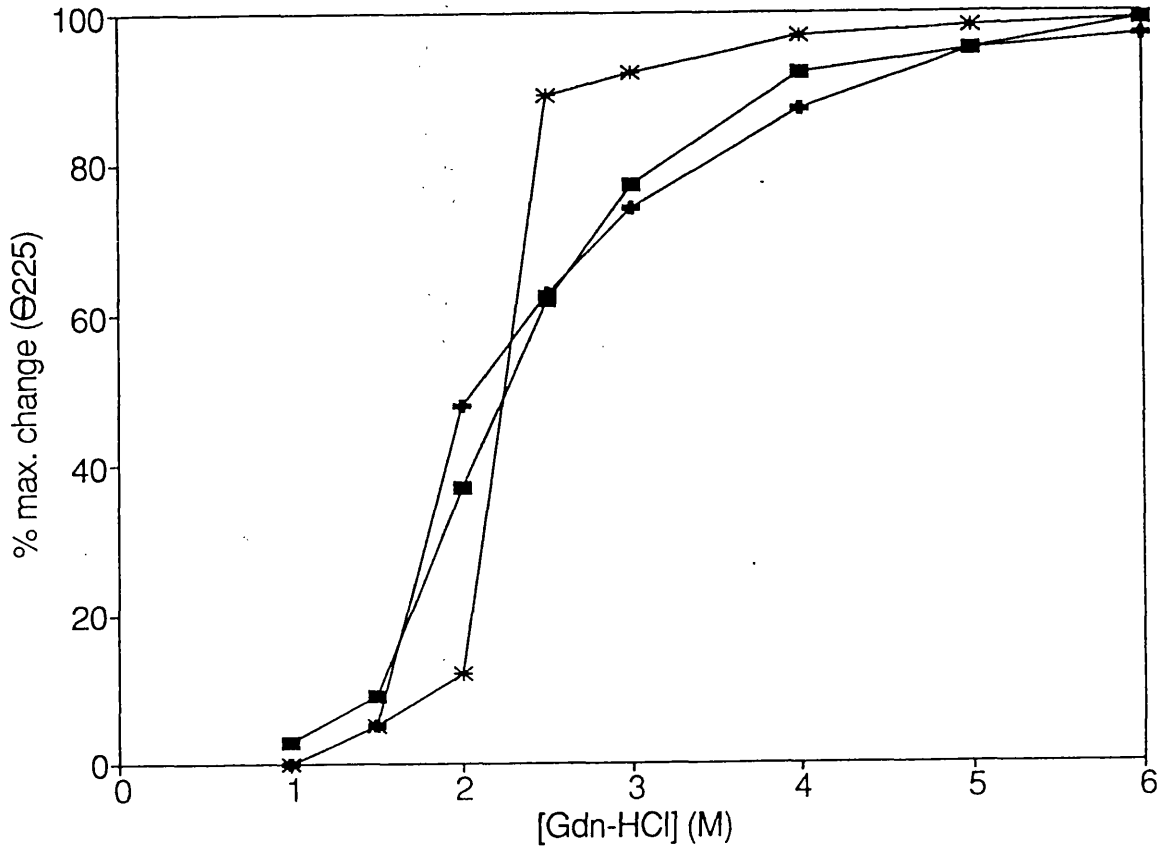


Figure 4.14: Gdn-HCl induced unfolding profile of pig heart citrate synthase at 20°C and *Tp.acidophilum* citrate synthase at both 20°C and 55°C, as judged by percentage changes in signal intensity at 222nm of circular dichroism spectra performed in the presence of known concentrations of Gdn-HCl.

- ■ Pig heart citrate synthase at 20°C
- ✱ *Tp.acidophilum* citrate synthase at 55°C
- ✱ *Tp.acidophilum* citrate synthase at 20°C

4.2.5 Guanidine-HCl inactivation of *Tp.acidophilum* citrate synthase

Samples of *Tp.acidophilum* citrate synthase were incubated at known Gdn-HCl concentrations and then assayed at 55°C to measure the enzymic activity present in the sample. The Gdn-HCl present in the incubation mixture will have been diluted out in the assay mixture, so effectively the assays were carried out in the absence of Gdn-HCl.

The %remaining activity was then plotted against Gdn-HCl concentration - see Fig.4.15. This procedure was repeated both at 20°C and 55°C.

4.3 DISCUSSION

In order to study the basis of thermal stability of proteins using site-specific mutagenesis experiments, the wild type of the protein must be fully characterised for comparative studies with the mutant. Thermal inactivation studies show that no enzymic activity is lost after incubation of the purified protein at 78°C for 10min, even though the optimum growth temperature of the organism is 55°C – 60°C. After incubation at 83°C for 10min no enzymic activity is remaining. Thus thermal inactivation occurs over a narrow temperature range.

Arrhenius plots of the thermal inactivation data will also be able to highlight the differences between wild type and mutant. If any mutations are made that decrease the thermal stability of the protein, the line on the Arrhenius plot will be shifted to the right along the x-axis and also the value of E_a for the mutant will highlight if the energy of activation of inactivation is altered in the mutant compared to the wild type.

Pig heart citrate synthase is dimeric and so will allow direct thermal stability comparisons with *Tp.acidophilum* citrate synthase. The data presented in Fig. 4.5 suggests that citrate synthase from pig heart is 80% inactivated after incubation for 10min at 35°C. This may be a function

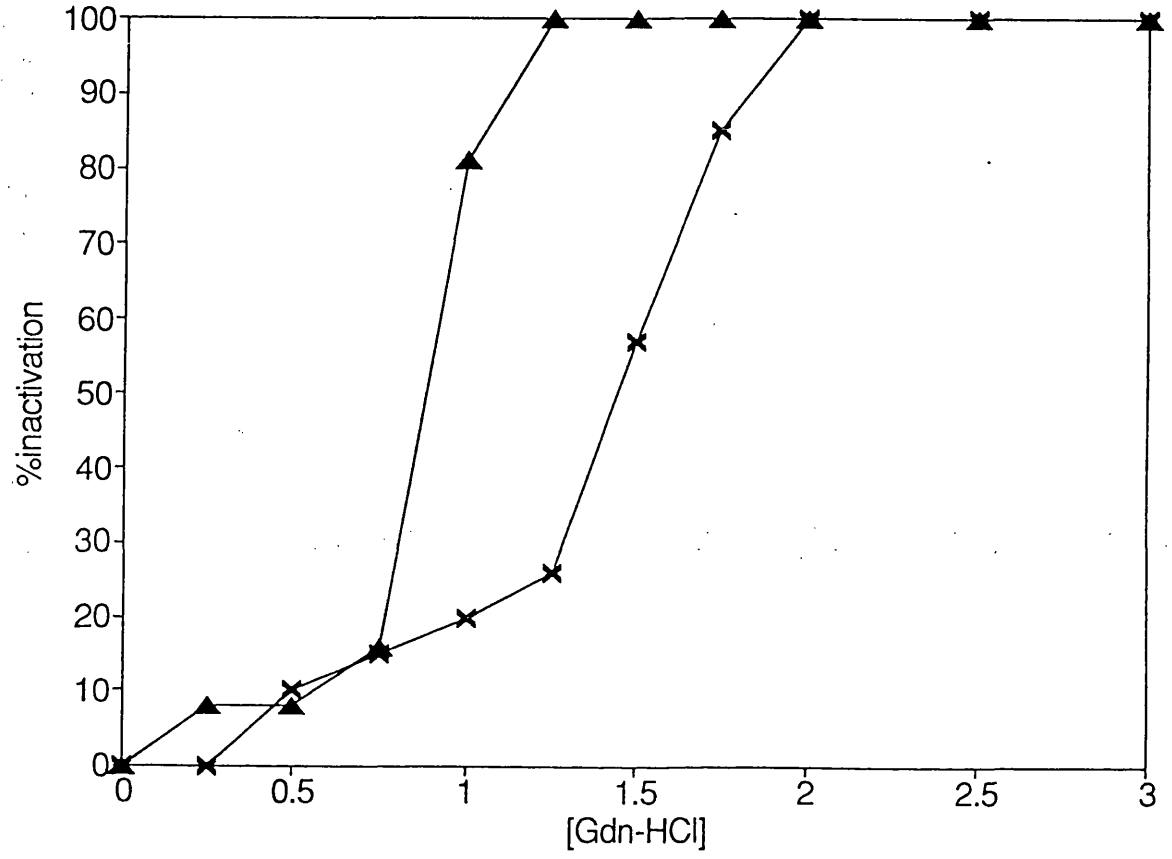


Figure 4.15: Gdn-HCl induced inactivation profile of *Tp.acidophilum* citrate synthase at $20°C$ and $55°C$. Samples were incubated at the appropriate temperature and assayed at $55°C$ for enzymic activity. The % activity lost was calculated with respect to sample incubated in the absence of Gdn-HCl.

▲ *Tp.acidophilum* citrate synthase at $55°C$

✖ *Tp.acidophilum* citrate synthase at $20°C$

of the assay conditions and protein concentration used. Arrhenius plots of their respective thermal inactivations give a very similar value for E_a, suggesting that the process required to disrupt the active site of the enzymes may be similar in the number and type of interactions involved in the thermal inactivation process. Fig.4.7 also shows the relative positions of the lines on their respective Arrhenius plots. Thus, any thermally destabilising mutant of *Tp.acidophilum* citrate synthase will have the position of its Arrhenius plot shifted towards the line for pig heart citrate synthase. Mutagenesis experiments will eventually be designed that will hopefully increase the thermal stability of *Tp.acidophilum* citrate synthase. The position of the line on the Arrhenius plot would then be shifted towards the position occupied by the *Sulfolobus solfataricus* citrate synthase.

Further characterisation of the stability of *Tp.acidophilum* citrate synthase was carried out using CD spectroscopy. A feature from Fig.4.8 is that the spectra are very similar at $25°C$ and $55°C$, indicating no change in structural conformation of the enzyme at the two temperatures. This is important as the crystallizations of *Tp.acidophilum* citrate synthase were carried out at $22°C$ and not at the organism's growth temperature of $55°C$.

The CD spectra for pig heart and *Tp.acidophilum* citrate synthases are very similar (results not shown) indicating a similar pattern of secondary structure for both enzymes. The predictions for α-helical content for pig heart citrate synthase give a slightly lower value than the actual crystallographic value but in the same range, whereas the predictions for β-sheet content are very high compared to the actual crystallographic value. Therefore the prediction for β-sheet content of *Tp.acidophilum* citrate synthase is probably far too high (see Chapter 9 for confirmation). The predicted value for α-helical content of *Tp.acidophilum* citrate synthase is very similar to

that of the predicted value pig heart citrate synthase, so the overall tertiary structural conformation of the two enzymes may indeed be the same despite the low sequence identity (see Chapter 9 for confirmation). Further evidence for this comes from the multiple sequence alignments showing the conservation of the catalytic residues throughout the citrate synthases, suggesting similar overall folds.

The CD spectra recorded at high temperatures show that possible large secondary structural changes are only seen after preincubation of *Tp.acidophilum* citrate synthase at 85°C for 20min, with the spectrum recorded at 85°C. The structure that generated this spectrum may not be indicative of a random coil, but may be caused in some part by a loss of total protein by precipitation during the incubation. The thermal unfolding profile (as judged by changes in ellipticity at 225nm) from the CD studies corresponds to the thermal inactivation profile (see section 4.2.1) although loss of activity does seem to precede loss of structure by a few degrees. This is expected since only a small change in the active site conformation would interfere with the catalytic activity of the enzyme, a change which would not be evident in the CD spectrum.

One feature evident from the 55°C denaturation spectrum is the increase in negative intensity signal in the presence of 1M Gdn-HCl. This seems to indicate that the native structure is in a tight conformation which is loosened slightly in the presence of a low concentration of Gdn-HCl. Enzymic activity is also increased by up to 50% if a low concentration of Gdn-HCl is present in the assay mixture, suggesting the enzyme may become more flexible in low Gdn-HCl concentrations, resulting in increased accessibility for the active site. Rehaber and Jaenicke (1992) observed larger levels of activation of the glyceraldehyde-3-phosphate dehydrogenase from the ther-

mophilic bacterium *Thermotoga maritima* at concentrations of Gdn-HCl up to 2M. This apparent tightness of the native enzyme may be related to the enzymes inherent thermal stability. *Tp.acidophilum* citrate synthase at 55°C and pig heart citrate synthase at 20°C have the same unfolding profile (followed by changes in ellipticity at 225nm), whereas the transition from folded to unfolded for *Tp.acidophilum* citrate synthase at 20°C is much sharper. Thus, at their respective working temperatures, the enzymes seem to unfold with the same profile, and therefore possibly following a similar Gdn-HCl induced unfolding pathway. However, the concentration of Gdn-HCl required to give 50%reduction in signal at 225nm (about 2M) is the same for *Tp.acidophilum* citrate synthase at both temperatures, and a slightly lower value (about 1.8M) for pig heart citrate synthase.

Investigations into whether the inactivation of *Tp.acidophilum* citrate synthase in Gdn-HCl is different at either 20°C or at 55°C have shown that a lower concentration of Gdn-HCl is needed to inactivate the enzyme fully at 55°C than at 20°C. The concentration of Gdn-HCl required to reduce the activity initially present by half is 0.75M at 55°C and 1.25M at 20°C. This may partly be accounted for by the proposed decreased flexibility of thermophilic enzymes at room temperatures compared to their working temperature (Jaenicke, 1991).

The thermal inactivation, Gdn-HCl inactivation, thermal denaturation and Gdn-HCl denaturation studies of wild type *Tp.acidophilum* citrate synthase should allow direct comparative studies between itself and mutants that have been designed to investigate the structural basis for the thermostability of this enzyme. Pig heart citrate synthase has also been characterised in a similar fashion which will allow mutants designed to increase its thermal stability to be fully compared with the wild type.

Chapter 5 : Crystallization, Data Collection and Analysis

5.1 X-RAY DIFFRACTION THEORY

5.1.1 Diffraction from a crystal

A crystal comprises a regularly repeating unit cell of axes lengths a, b and c and inter-axial angles α, β and γ. Unit cells can only belong to seven distinct crystal systems with a possible 14 crystal lattices which can be further sub-divided into 32 crystal classes according to the arrangement of the symmetry elements with respect to each other. Proteins can only crystallize in enantiomorphic space groups reducing the number of possible space groups to 65.

5.1.2 Bragg's Law

Bragg expressed the conditions necessary to produce diffraction from a series of parallel planes of atoms as present in a crystal as::

$$n\lambda = 2d \sin\theta$$

where n is an integer, λ is the wavelength of the incident beam, d is the interplanar spacing and θ is the angle of incidence and reflection. Therefore diffraction is only seen if the angle θ satisfies the above equation.

5.1.3 The Laue equations

An alternative condition for diffraction by a crystal can given by the Laue equation:

$$\mathrm{a}.\mathbf{S} = \mathrm{h}$$

$$\mathrm{b}.\mathbf{S} = \mathrm{k}$$

$$\mathrm{c}.\mathbf{S} = \mathrm{l}$$

where h, k and l are integers. **S** is the diffraction vector and is defined by s-s_0, where s_0 represents the direction of the incident beam and s the direction of the diffracted beam.

5.1.4 Calculation of Structure factors and Electron density

Any wave can be described by means of a Fourier series as the sum of simple waves. A three-dimensional wave consists of three frequencies in the direction x, y, and z and therefore must be described by three variables h, k, and l. A general Fourier series for the wave f(x,y,z) has the form:

$$f(x,y,z) = \sum_h \sum_k \sum_l F_{hkl} \exp(2\pi i(hx_j + ky_j + lz_j))$$

The diffraction by a molecule in a crystal lattice is given by the structure factor equation. The structure factor for reflection *hkl* can be expressed in form of a Fourier series in which each term gives the contribution of one atom to reflection hkl:

$$\mathbf{F}(hkl) = \sum_{j=1}^{n} f_j \exp(2\pi i(hx_j + ky_j + lz_j))$$

where f_j is the atomic scattering factor of an atom which is treated as a simple sphere of electron density.

$\mathbf{F}(hkl)$ possesses amplitude, frequency and phase. The frequency is that of the x-ray source and the amplitude is proportional to the square root of $I(hkl)$, the intensities recorded during data collection.

$\mathbf{F}(hkl)$ can also be written as the sum of contributions of the electron density, $\rho(x,y,z)$, from each point within the unit cell volume,V :

$$\mathbf{F}(hkl) = \int_V \rho(x,y,z) \exp(2\pi i(hx + ky + lz))dV$$

ie. $\rho(x,y,z)$ is the Fourier transform of $\mathbf{F}(hkl)$

This operation is reversible resulting in the electron density equation:

$$\rho(x,y,z) = 1/V \sum_h \sum_k \sum_l \mathbf{F}(hkl) \exp(-2\pi i(hx + ky + lz))$$

Therefore the calculation of an electron density map requires information concerning the phase $\alpha(hkl)$ must be known. This is a primary problem in the elucidation of a protein structure from x-ray diffraction data.

The phase problem in protein crystallography can be overcome in one of two basic ways - heavy atom isomorphous replacement and/or molecular replacement. The theory behind the latter will be discussed in detail in Chapter 6, but no detailed theory will be given for the former due to the technique only being employed at its preliminary stages.

5.2 CRYSTALLIZATION INTRODUCTION

A major cause of heartache and anxiety in the determination of the 3-dimensional structure of a protein is obtaining crystals deemed suitable for x-ray diffraction data collection. The ability to rationally design suitable conditions for protein crystal growth is still in its infancy, although attempts have been made including the incomplete factorial approach (Carter *et al.*, 1988) and the 'Magic 50' (Jancarik and Kim, 1991). Therefore a more systematic approach must be used in which individual parameters must be explored through logical refinement and inspiration. Crystals of a compound grow when a solution of the compound is brought into a supersaturated state, eg. by gradual solvent evaporation, and its return to the equilibrated saturated solution causes a shift of the solute molecules into the solid state, either crystalline or amorphous. The solid state will only occur if a stable nucleus has formed in the supersaturated solution, after which the stable nucleus will grow until equilibrium has once again been reached. Therefore in order to crystallize a compound, conditions must be found in which a supersaturated state is formed followed by spontaneous stable nucleus formation.

Proteins are highly complex physical-chemical systems whose intrinsic properties are affected by many external influences such as temperature, pH, ionic strength and solvent composition. Microhomogeneity of the protein sample is a key factor in protein crystallography and so every effort must be made to ensure protein purity and constant environmental conditions.

5.3 RESULTS

5.3.1 Crystallization trials

The crystallization trials were performed using the hanging drop vapour diffusion method (McPherson, 1990). The hanging drops were placed on siliconised glass cover slips and equilibrated over 1ml of precipitant solution in 24-well tissue culture plates. The seal was made by a layer of glisseal grease. Initial trials involved the use of either $(NH_4)_2SO_4$, PEG3350 or PEG6000 as the precipitating agent, using a protein concentration of 15mg/ml in a 4μl hanging drop. A range of precipitant concentrations and pH values were used with specific buffers, eg Tris-HCl, and trials repeated at 25°C and 4°C. Small hedgehog crystals were seen to grow from a solution containing 10.5% (w/v) PEG3350 (pH8.0) after 2–3 weeks at 25°C. Further trials were then carried out using a finer grid based on these conditions, using 20mM Tris-HCl as the standard buffering agent. Crystals (Type1) suitable for X-ray diffraction studies were grown from a solution containing 9% (w/v) PEG3350 (pH8.2) at a protein concentration of 15mg/ml. Crystals grew to maximum dimensions of 0.2mm x 0.3mm x 0.6mm in 3–4 weeks at 25°C. A second crystal form (Type2), as judged by subsequent analysis (see section 5.5.4), was also grown using 9% (w/v) PEG3350 (pH8.2) but at 10mg/ml protein concentration and with a drop of toluene added to the reservoir. Crystals grew to maximum dimensions of 0.3mm x 0.4mm x 1.0mm in 3–4 weeks at 25°C - see Fig. 5.1 (The presence of fly in the protein sample did not have a deleterious effect on reproducibility of growth of this crystal type). A third crystal form (Type3), again as judged by subsequent analysis, was grown using 8.5% (w/v) PEG3350 (pH8.2) at 10mg/ml protein concentration and with a drop of toluene added to the reservoir. Crystals grew to maximum dimensions of 0.3mm x 0.4mm x 1.5mm in 3–4

Figure 5.1: Type2 crystals grown from 9% (w/v) PEG3350 (pH8.2) at a protein concentration of 10mg/ml, with a drop of toluene in the reservoir. The crystals dimensions are 0.3mm x 0.4mm x 1.0mm

months at $25°C$.

Reproducibility of crystal quality was a problem throughout the trials. The same conditions did not always give the same quality of crystal each time.

Crystallization trials were also set up with substrates present in the hanging drop. The substrate(s) of choice were preincubated with the native protein of a known concentration for 1hr at $55°C$ before addition of the precipitant in the hanging drop. Crystals with 5mM citrate present were grown from a solution containing 10.5% PEG3350 (pH8.2) at a protein concentration of 5mg/ml. Crystals grew to maximum dimensions of 0.1mm x 0.1mm x 0.3mm in 2–3 days. Crystals with 2.5mM oxaloacetate present were grown from a solution containing 10% PEG3350 (pH8.2) at a protein concentration of 7.5mg/ml and grew to maximum dimensions of 0.2mm x 0.2mm x 0.4mm in 1–2 days. Crystals with 5mM citrate and 2.5mM CoA present were grown from a solution containing 8.5% PEG3350 (pH8.2) at a protein concentration of 10mg/ml which grew to maximum dimensions of 0.2mm x 0.2mm x 0.8mm in 2–3 weeks.

5.4 CRYSTALLIZATION DISCUSSION

The growth of crystals of native *Tp.acidophilum* citrate synthase suitable for x-ray diffraction studies has proved to be long and problematic. The appearance of small hedgehogs from the preliminary crystallization trials indicated that the purification procedure had yielded protein of sufficient purity for crystallization. After minor alterations of the 'hedgehog' conditions, larger crystals grew but were very fragile and tended to be stacked. To overcome this problem of crystal stacking toluene was added to the reservoir in various conditions. Toluene was hoped to cause a slight change in the microenvironment of the hanging drop sufficient enough to

promote isolated crystal formation. Further trials in the presence and absence of toluene yielded distinct crystals of a suitable size for diffraction studies.

Since high resolution data is required for informative analysis of the x-ray structure of a protein, crystallization trials were continued in the search for suitable crystals. These trials eventually led to the discovery of three different crystal forms of the enzyme.

The phenomenon of the same protein crystallizing in multiple space groups has been seen previously with proteins such as lysozyme, insulin and pepsin. Lysozyme crystallizes in a tetragonal form in the presence of NaCl at an acidic pH and at a temperature lower than $25°C$ but in an orthorhombic form at higher temperatures (McKenzie and White, 1991 and references therein). Temperature is just one of many external factors which can affect crystal growth, with the resulting effect of these changes causing there to be a change in the protein-protein interactions within the crystal leading to the formation of a different crystal lattice. The only factors that were altered in the formation of Type 1 and Type 2 crystals were the reduction in protein concentration and the addition of toluene in the reservoir of the Type 2 crystal. Thus either one or the combination of the two must somehow be affecting how the protein crystallizes. Another factor which could account the variation is that the protein from which Type 1 and Type 2 crystals grew was purified in different batches, although reproducibility of crystal forms has been achieved. The above factors may also be an explanation of the problems in reproducing crystal quality.

On binding of substrates to pig heart citrate synthase the small domain rotates 18° with respect to the large domain, forming the 'closed' form of the enzyme. Similar conditions were found to be necessary to crystallize

Tp.acidophilum citrate synthase in the presence of its substrates than in their absence. Crystals formed in the presence of 5mM citrate after a few days, which is more rapid than the native protein. This may be due to the 'closed' form being able to pack more favourably into a crystal lattice than the 'open' form. Unfortunately after a few days the crystals growth ceased, with the crystal dimensions no longer than 0.3mm. A crystal was mounted in a quartz capillary but the diffraction seen was very weak and no data collection/analysis was carried out. No larger crystals could be formed.

Crystals also formed within 1-2 days in the presence of 2.5mM OAA. These crystals were slightly larger than those grown in the presence of citrate, but after further 2 days the crystals tended to crack and disintegrate. OAA is unstable and degrades fully after a day at room temperature. Crystal formation is rapid whilst OAA is present but after the OAA degrades the crystal lattice may be disrupted by the equilibrium between the 'open' and 'closed' forms of the enzyme, leading to crystal cracking and eventual disintegration. No data have been collected on these crystals due to their instability, but further investigation into this crystal form could be carried out by using a stable analogue of OAA, such as flouro-OAA to try and form a stable crystal.

The most successful crystallization of *Tp.acidophilum* citrate synthase with substrates was using 1hr soaks of CoA (2.5mM) and citrate (5mM). This complex formed stable crystals which continued to grow for 2-3 weeks, after which time they were large enough for x-ray diffraction studies. The conditions for crystal growth were identical to that for growing Type 3 native crystals but without the toluene in the reservoir.

5.4.1 Differences between crystals grown from Red GelA and chromatofocussing purifications

The majority of the crystallization trials were carried out using protein purified using chromatofocussing, this procedure being the first one developed. Crystals have been grown using protein purified using Matrex Red GelA but they tend be much smaller than crystals grown under the same conditions using protein purified using chromatofocussing. This is probably due to the different levels of protein purity obtained from the two procedures. All the data detailed below were collected using crystals grown from protein purified using chromatofocussing. Thus far only data for Type 2 and 3 crystals have been collected using protein purified using matrex Gel RedA; this data being part of the search for isomorphous heavy atom derivatives - see Chapter 7.

5.5 DATA COLLECTION AND ANALYSIS

5.5.1 The Area detector

All data were collected on the in-house Siemens X-100 area detector mounted on a Siemens rotating anode emitting CuK_α radiation of wavelength 1.54Å, the latter being selected by a graphite monochromator. The detector is mounted on a moveable 2θ arm whose axis is coincident with the rotation ω axis of a 3-circle goniostat, which are controlled by a PC. The x-ray photons produced cause ionization of the xenon gas which is present at 4atm in the detection chamber. The high pressure causes the ionization events to occur just behind the 1mm concave beryllium window. The electrons produced are then accelerated towards the anode wires causing secondary ionization events resulting in amplification of the original signal. A set of 2 parallel planes of wires (horizontal and vertical) in the cathode allows the 2-dimensional position of the ionization events to be determined by a position decoder. The data are stored as a 512x512 pixel array which is written to disk at the end of each exposure and then transferred to a microVax 4000 via an ethernet linkage. A pixel with a count greater than 255 is classed as an overflow.

Local heterogeneity within the detector caused by irregularities in the cathode wire spacing is eliminated by a 'flood field' correction using a ^{55}Fe x-ray source at the working distance from the crystal, ensuring a smooth response over the whole of the detector surface. Geometric distortion is corrected by recording an image of a brass plate (a brass shield with holes drilled in it at precise intervals) using an ^{55}Fe x-ray source at the working distance from the crystal. The observed pattern of spots is then compared to the known pattern of spots that should be generated from the brass plate and a 2-D distortion correction is applied to each frame of data processing.

5.5.2 Data collection

A crystal was drawn into a thin-walled quartz capillary tube of suitable diameter for the size of the crystal, the most of the mother liquor drained from around the crystal and the ends of the tube sealed with beeswax. The capillary tube was mounted on the goniometer head with plasticine.

Once suitable crystals for diffraction studies have been obtained, the data collection procedure must be at an optimum to ensure that significant data of the highest possible resolution are collected. The specific starting setup of the area detector and its controls are dependent on the parameters detailed below.

The minimum distance in cm from the crystal to the detector (XTDD) needed to resolve adjacent spots can be approximated to be the length of the largest primitive cell axis (in Å) divided by 10 (Taylor, G. L., personal comm.). When the unit cell dimensions of the crystal are unknown the distance must be relatively large, eg.30cm, to avoid spot overlap. But on determination of the cell dimensions the XTDD can be reduced to facilitate higher resolution data to be collected. When the 2θ arm is zero the highest resolution data will not be collected, but the detector can be rotated around the 2θ axis in order to record higher resolution data. The choice of the 2θ swing is dependent on the strength of the diffraction. A relationship exists between XTDD and the 2θ swing which defines the maximum resolution data that can be collected - see Fig. 5.2. The 2θ setting for *Tp.acidophilum* citrate synthase crystals has ranged from 0°–20° depending on crystal size and its diffractive power. The XTDD for type 1 and 2 crystals was 13cm, although for type 3 crystals a XTDD of 15cm was necessary to resolve adjacent spots, using a 0.3mm collimator.

The need for collection of the largest number of significant reflections

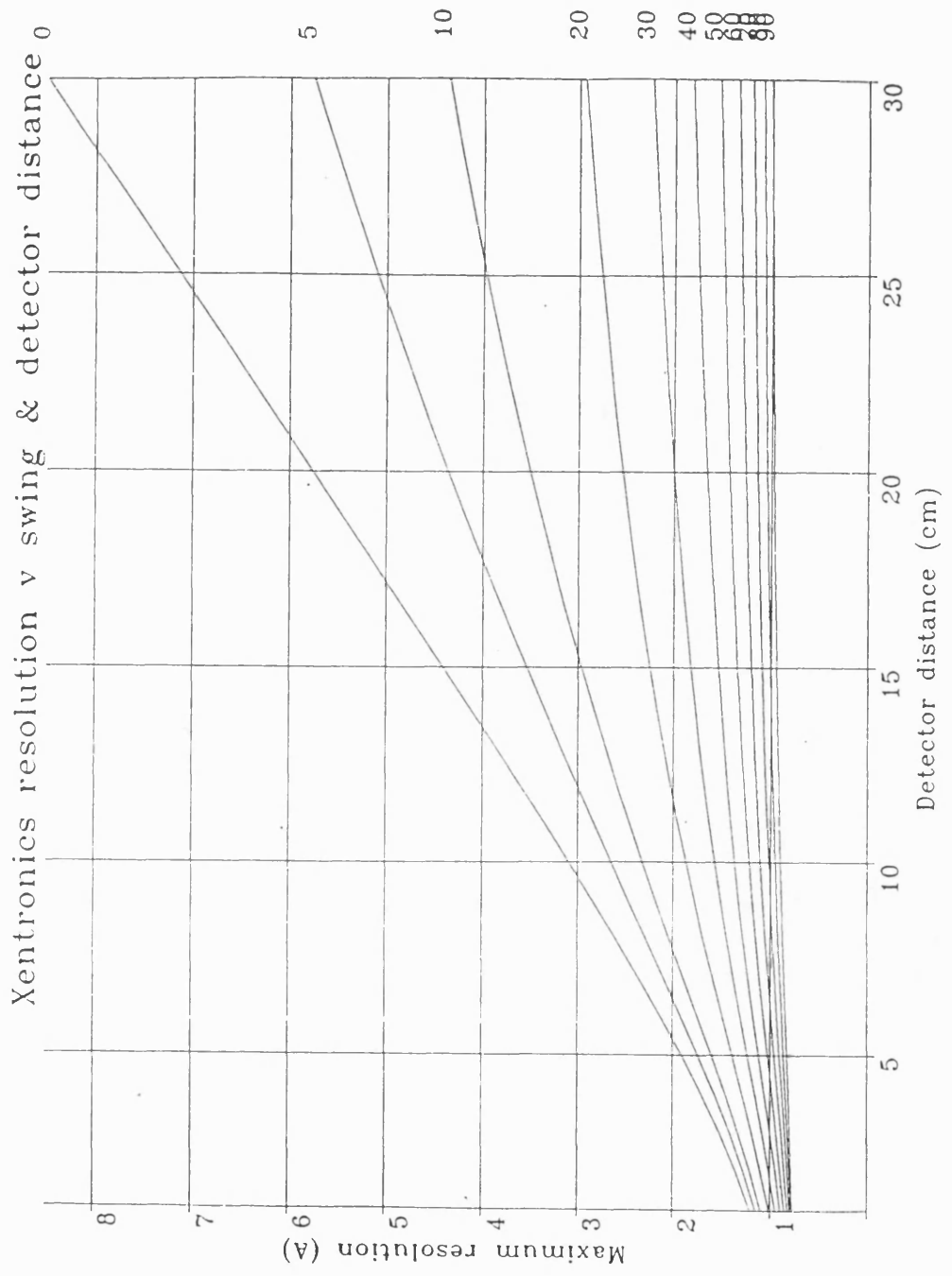


Figure 5.2: Relationship between the crystal to detector distance and the value of 2θ swing, with respect to maximum possible resolution that can be collected.

possible has to be counterbalanced by the time available for data collection and any deterioration of the crystal caused by X-rays; the latter judged by the number of overflows and visual interpretation of sequential frames, using the Siemens Area Detector Integrated environment program (SADIE). Therefore the length of time of exposure for each frame must be a balance between these factors. The intensity of the diffraction pattern is directly proportional to the volume of the crystal and indirectly proportional to the volume of the unit cell. The larger the overall dimensions of the crystal, the stronger the diffraction that tends to be seen and thus the shorter the exposure time required (although many other parameters , eg. solvent content of the crystal, unit cell dimensions and collimator size, can affect the strength of the diffraction seen). The exposure times for *Tp.acidophilum* citrate synthase crystals has ranged from 120sec–300sec per 0.25° frame depending on the crystal dimensions, resulting on average in 30 or more overflows per frame. The length of exposure is also dependent on the power of the X-ray generator; the standard setting was 45kV and 80mA.

The amount of diffraction data required to record every reflection possible from a crystal depends on the space group of the crystal. The higher the symmetry the fewer the degrees of ω rotation needed to record every reflection. A crystal belonging to the monoclinic space system (eg. crystal type2) needs 180° rotation to theoretically record every reflection at least once, whereas a crystal belonging the orthorhombic crystal system (eg. crystal types 1 and 3) only needs 90° rotation due to the extra symmetry present in the crystal. In order to reduce the significance of errors in the data collection and processing, multiple copies of each reflection must be collected, resulting in the practical need of collecting more than the theoretical number of frames.

Data were collected as a series of contiguous non-overlapping frames; the crystal being rotated around ω, with each frame representing a 0.25° oscillation.

All the data presented below were collected at room temperature and from single crystals for each crystal type.

5.5.3 Data processing

All data were processed using the XDS suite of programs (Version 2) (Kabsch, 1988). The process is a fully automated process which reads in raw Xentronics frames and produces a list of scaled hkl intensities. A flow diagram of the order of the individual programs is given in Fig. 5.3, with details of the individual programs given below.

XYCORR determines the distortion of the diffraction pattern in X and Y, arising from the projection of ionization events occurring near the concave beryllium/gas interface onto the planar recording wires. The recently recorded brass plate is read and an allowance is calculated for the geometric distortion by calculating the difference between the observed and known positions of the spots.

INIT reads the first 30 frames and calculates the average background count. At a given pixel position if the count is $<3\sigma$ away from the lowest count the pixel is 'accepted', with the mean of the 'accepted' count being the background count. Active pixels are defined as having >20% of the mean count.

COLSPOT collects spots for autoindexing and finds spots above the background count, using the pattern recognition procedure described above. Equivalent spots on adjacent frames are merged. A list (SPOT.XDS) of observed spot coordinates and their spindle positions are produced.

IDXREF reads in SPOT.XDS and predicts the unit cell of the crys-

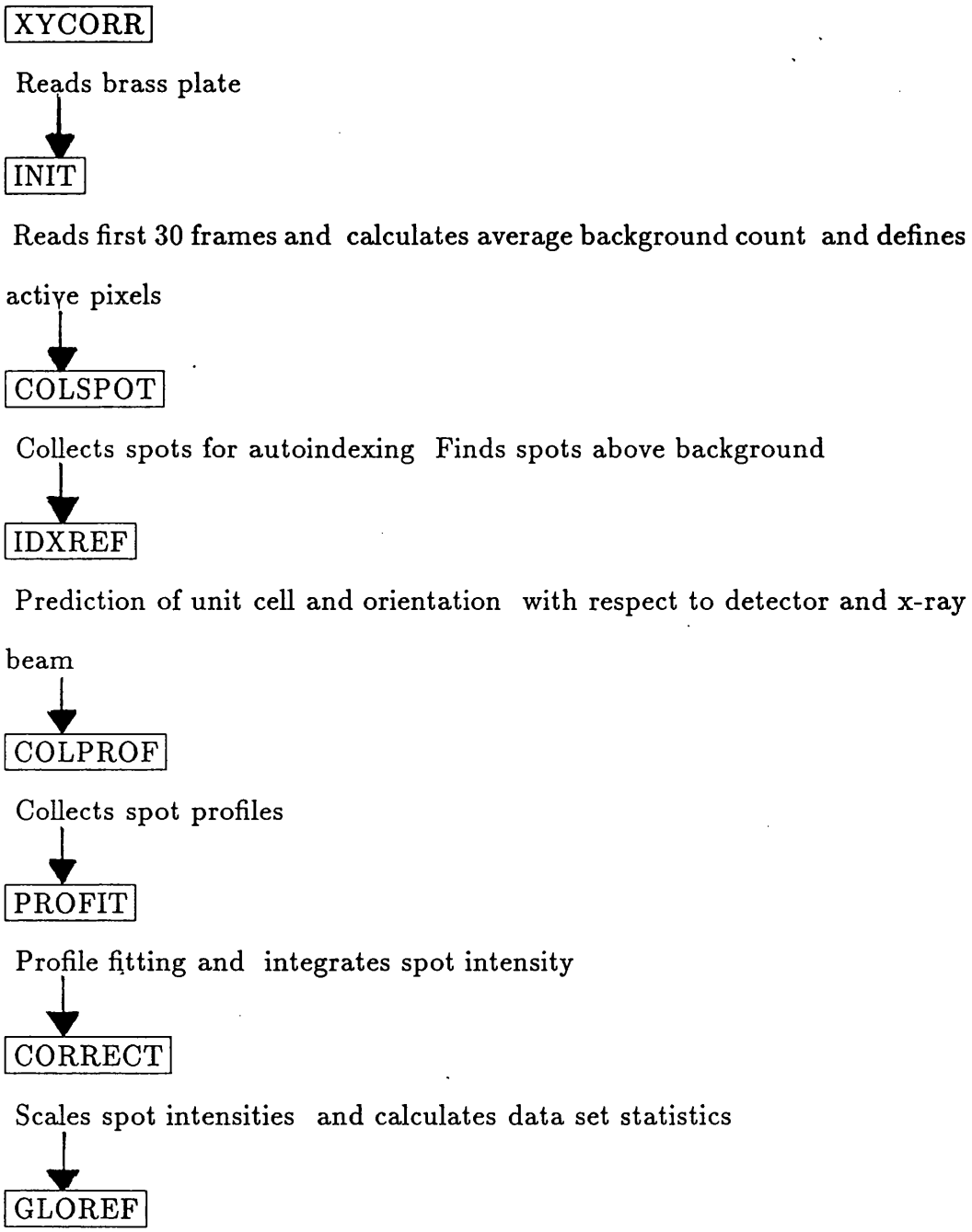


Figure 5.3: Flow chart representing the individual steps in XDS (Kabsch, 1988) for processing of x-ray diffraction data.

tal and its orientation with respect to the detector and x-ray beam via autoindexing of the spots.

COLPROF collects all the spot profiles. All predicted spots are scanned and each is numerically labelled by the indices of the nearest reflection to which it could belong. The indices are checked for whether they obey the general conditions limiting the possible reflections for the space group of the crystal. If they are not satisfied, the nearest reflection to which the pixel belongs, is classed as being systematically absent, given a negative label and is used to update the background count. The 3-dimensional reflection profile is represented by a box of 9x9x9 grid points. The step sizes between the grid points are derived from values for beam divergence and mosaic spread. All the profiles tend to be very similar to each other since all spots are treated as if it were on the Ewald sphere and had followed the shortest path through it.

Profile fitting is implemented in PROFIT, which integrates the spot intensity. Integration must be able to distinguish between signal and background points within each reflection box. Therefore weak and strong signals have the same normalised profile. The reflection intensities are estimated by a 2-pass integration procedure. Firstly, the reference profiles are learned as a function of position on the detector and spindle angle. If a grid point within the average profile box is >2% of the peak maximum then it is classed as a signal. Secondly, the intensities for each reflection are estimated using the nearest reference profile.

CORRECT ensures that the intensities of the spots are on the same scale, in order to correct for crystal decay and variation in detector sensitivity with a continuous correction factor. After correction the integrated intensities of all the reflections are sorted in order of increasing key values

derived from their unique reflection indices and output as a list of intensities of the reflections.

GLOREF refines the unit cell parameters and in a more statistically relevant manner than IDXREF due to the use of all the reflections in the calculation.

XSCALE (Kabsch 1988) was used to apply an empirical absorption correction to individual intensities of the data set. Multiple data sets of the same crystal type can also be scaled and merged into a single list of intensities by XSCALE. The scaling of heavy atom soaked crystals was carried out by XX (a slightly modified version of XSCALE (G.L.Taylor)) which preserves the intensities of Friedel pairs (I+, I-). The list of scaled intensities were then converted to a list of structure factors using X2L (the CCP4 program OXMAKE modified by Dr. G. L. Taylor).

5.5.4 Space group determination

XDS allowed the rapid (after 30 frames of data have been collected) and usually straightforward determination of the cell dimensions of the crystals. Types 1 and 2 proved to be relatively simple to assign specific space groups, but type 3 was more problematic.

Type 1 crystals were the first of the crystal types to have data collected from and belong to the space group $P222_1$ (orthorhombic crystal system) , as judged by systematically absent or weak ($F < 3.5\sigma$) 00l reflections, diffracting to a resolution limit of 3.2 Å. The unit cell parameters are a=80.9 Å, b=103.8 Å, c=98.3 Å.

The Type 2 crystals belong to space group $P2_1$ (monoclinic crystal system) , as judged by systematically absent or weak ($F < 3.0\sigma$) 0k0 reflections, diffracting to a resolution limit of 2.4 Å. The unit cell parameters are a=53.8Å, b=173.8 Å, c=86.7 Å and $\beta = 97.1°$.

Type 3 crystals diffracted to a resolution limit of 3.0 Å but preliminary native data analysis could not specifically assign the space group but either belong to space group P2, $P2_1$ or $P2_12_12_1$, with initial space group predictions by IDXREF. The amount of data collected was not enough to resolve this conflict. Data processing statistics were of similar quality in both orthorhombic and monoclinic crystal systems but the number of misfits rejected by CORRECT was far greater when processed in the monoclinic system. The specific assignment to the orthorhombic crystal system was further strengthened upon more data being collected for this crystal type whilst undertaking heavy atom soaking experiments. Therefore crystal type 3 most probably belongs to space group $P2_12_12_1$ with unit cell parameters of a=51.5 Å, b=112.5 Å, c=123.1 Å.

The determination of cell dimensions for a crystal allows the assignment of the number of molecules likely to be present within the asymmetric unit. Matthews (1968) reported on the range of solvent content that is likely to occur in protein crystals and inferring the number of molecules within the asymmetric unit. Assuming a relative molecular mass of 43 000 Da for the monomer of *Tp.acidophilum* citrate synthase (Sutherland *et al.*, 1990), values for V_M and solvent content that lie in the range predicted by Matthews, revealed that type 1 and 2 crystals to contain one dimer and two dimers within the asymmetric unit respectively. Crystal type 3 probably contains one dimer within the asymmetric unit (35% solvent content) although protein crystals with 95% solvent content (tropomyosin) have been reported (Matthews, 1968).

Type1 V_M = 2.58 Å^3/Da 40% solvent One dimer per asymm. unit

Type2 V_M = 2.53 Å^3/Da 40% solvent Two dimers per asymm. unit

Type3 V_M = 2.20 Å^3/Da 35% solvent One dimer per asymm. unit

The confidence of the prediction of the number of molecules in the asymmetric unit is of great importance if a molecular replacement strategy is to employed for solving the ubiquitous phase problem in protein structure elucidation.

5.5.5 Data processing statistics

All relevant data processing statistics (for all data) are given in Table 5.1. The data sets for crystal types 1 and 3 were collected in a continuous series of non-overlapping frames and therefore could be fully processed in one stage, but the final data set for crystal type 2 was comprised from three individual data collection procedures. The level of quality of the final merged data set for crystal type 2 was superior to the other two crystal types, and therefore will be discussed in more detail. Each one was processed separately, merged and then scaled together, using XSCALE. A critical factor in the success of the merging is the isomorphous nature of individual data sets; non-isomorphism manifests itself in an increase in the R_{merge} of the merged data set compared to the value for the individual data sets. The cell dimensions for the three individual data sets are given in Table 5.2. Although the three data sets do not show identical cell dimensions, the change between them is not greater than 0.2% in any of the axes and in β. The correlation (between common intensities from individual data sets after a B-factor correction) was greater than 0.98 between the three data sets. Upon the merging the three data sets together the R_{merge} rises as higher resolution data shells are included in the analysis, while maintaining at a reasonably low level to 2.8 Å- Table 5.2. Nevertheless, this trend was seen in each of the individual data sets. Therefore the scaling of individual data sets into a single merged data set did not prove to be significantly detrimental, due to the increased redundancy of data.

Native data	Space group	Number of observations	Unique reflections	Resolution (Å)	% Completeness	overall
					$>0\sigma$	Rsymm
TypeI	$P222_1$	20075	9394	3.6	79.3	11.7

Cell dimensions	a	b	c	α	β	γ
	80.50	98.50	104.40	90.0	90.0	90.0

Native data	Space group	Number of observations	Unique reflections	Resolution (Å)	% Completeness	overall
TypeII					$>0\sigma$	Rsymm
csnat1	$P2_1$	36475	22558	3.0	56.4	3.1
csnat3	$P2_1$	52841	42055	2.5	52.9	2.7
csnat4	$P2_1$	40578	27168	2.5	34.7	3.8
xscaled csnat134	$P2_1$	129680	52490	2.5	73.9	6.4 (Rmerge)

Cell dimensions	a	b	c	α	β	γ
	53.70	173.70	86.70	90.0	97.1	90.0

Native data	Space group	Number of observations	Unique reflections	Resolution (Å)	% Completeness	overall
					$>0\sigma$	Rsymm
TypeIII	$P2_12_12_1$	18298	8756	3.6	72.7	8.8

Cell dimensions	a	b	c	α	β	γ
	51.55	112.22	122.74	90.0	90.0	90.0

Table 5.1: Data processing statistics for crystal types 1, 2 and 3 of *Tp.acidophilum* citrate synthase. All data were collected from single crystals. $Rsymm = \sum \mid I - \langle I \rangle \mid / (\sum \langle I \rangle)$. Rmerge is the equivalent value for merged data sets.

Table 5.2: Data processing statistics for different resolution shells for the merged data set of crystal type2.

HIGH RESOLUTION LIMIT	NUMBER OF OBSERVED REFLECTIONS	NUMBER OF UNIQUE REFLECTIONS	R-*MERGE* (%)	COMPLETENESS OF DATA (>1σ)
6.00	17430	3969	4.6	99.2%
4.50	21675	5435	6.2	99.5%
3.60	30178	8848	7.1	99.1%
3.20	19045	7239	8.3	91.6%
3.00	8720	4923	7.0	85.1%
2.80	10641	6330	8.4	82.7%
2.70	5945	3811	9.6	79.2%
2.60	6210	4270	11.3	73.7%
2.50	5933	4451	13.1	64.4%
2.40	3903	3214	14.6	39.0%
total	129680	52490	6.4	

Cell dimensions	a	b	c	α	β	γ
csnat1	53.772	173.756	86.746	90.0	97.12	90.0
csnat3	53.622	173.669	86.683	90.0	97.23	90.0
csnat4	53.603	173.495	86.598	90.0	97.38	90.0

5.6 DATA ANALYSIS OF SUBSTRATE BOUND CRYSTALS

The data collection/analysis protocol was the same as described above. The strength and limit of the diffraction seen for citrate bound crystals did not warrant the collection of a data set, whereas for the citrate + CoA bound crystals diffracted strongly and to a limit of 3.2Å. The unit cell dimensions and data processing statistics are presented in Table 5.3. The unit cell dimensions and space group assignment are identical to those of crystal type 1 and therefore suggests that the substrates in this case may not have bound to the protein since a large conformation change is known to occur upon binding of substrates to pig heart citrate synthase, which would probably cause a change in the cell dimensions of the crystals. Although, efforts to merge this and crystal type 1 data sets together produced very low correlations, in the order of 0.5.

5.7 DATA COLLECTION/ANALYSIS DISCUSSION

The in-house area detector has proved to be an invaluable tool in the collection of crystallographic data for *Tp.acidophilum* citrate synthase. The rapid detection of crystal imperfections (eg. twinning) and the rapid determination of space groups (XDS) has allowed many crystals to be analysed. Data from crystal type 2 proved to be superior in both the maximum resolution of diffraction and the processing statistics, even though the large b-axis required a longer XTDD than the other crystal types. The latter was counterbalanced by the relative size of the crystal compared to the other types, allowing a greater 2θ angle to be employed. Once the detector parameters had been optimized to ensure collection of the highest possible resolution data, XDS has allowed the rapid processing of the data. The preliminary twinning of crystals caused many problems in the collection of suitable data for structural elucidation. Many twinned (with hindsight)

Table 5.3: Data processing statistics for substrate (citrate + CoA) bound crystals of *Tp.acidophilum* citrate synthase.

Native data	Space group	Number of observations	Unique reflections	Dmax	% Completeness	overall Rmerge
TypeI	P2221	21821	13014	3.2	81.3	5.4

Cell dimensions	a	b	c	α	β	γ
	80.88	98.33	103.82	90.0	90.0	90.0

crystals were mounted and data were collected from them, but data processing failed to resolve space group determination and were therefore of no use. A problem encountered many times was the slippage of the crystal within the capillary tube causing the X-ray beam to be incident with different parts of the crystal in the same data collection. In many cases this caused irredeemable problems with subsequent data analysis, but in some instances the problem was alleviated by processing of the data in small distinct sections (although, such processing did not contribute to the aforementioned data sets). Potential crystal damage by X-ray radiation was minimized (where possible) by successive translations of the crystal with respect to the X-ray beam. In this way, crystal type 2 proved to be extremely resilient to X-ray based deterioration, with over 2000 frames of data collected from a single crystal.

Crystal parameters of open and closed forms of pig heart citrate synthase

The open form of pig heart citrate synthase crystallizes in space group $P4_12_12$ with cell dimensions a=b=77.4Å and c=196.4Å. The closed form crystallizes in space group C2 with cell dimensions a=104.14Å , b=78.25Å c=58.40Å and β=78.5°.

Chapter 6 : Molecular Replacement

6.1 MOLECULAR REPLACEMENT THEORY

The molecular replacement technique is used in two basic strategies for gaining phase information. The first kind is in the search for the orientation of non-crystallographic symmetry elements within the asymmetric unit. Electron density maps have been improved by 'averaging' around local symmetry elements; this technique has been invaluable in the determination of virus structures (Luo *et al.*, 1987 and Acharya *et al.*, 1989).

The second kind which will be discussed in more detail here is in the derivation of phase information for an unknown structure from a known structure. The superposition of the known structure onto the unknown structure via a six-dimensional search is commonly split up into two distinct parts, due to limitations in computing power. Firstly the determination of the correct orientation of the known structure in the cell of the unknown (the rotation function), followed by the positioning of the correctly orientated model in the cell of the unknown (the translation function).

6.1.1 Rotation functions

The corner stone of the rotation function is the properties of the Patterson function (Patterson, 1934):

$$\mathrm{P}(\mathbf{u}) = \int_{\mathrm{v}} \rho(\mathbf{x}).\rho(\mathbf{x}+\mathbf{u})\,dv$$

where P($\mathbf{u}$) is the value of the function at a particular point $\mathbf{u}$(u,v,w), which is a product of of two values of ρ at position x and x+$\mathbf{u}$ and integrated over the whole unit cell. P($\mathbf{u}$) has significant value when the vector $\mathbf{u}$ corresponds to an interatomic vector.

The Patterson function may also be expressed in terms of structure

factor amplitudes:

$$P(\mathbf{u}) = 1/V \sum_{h} F(\mathbf{h})^2 \exp(-2\pi i \mathbf{h}.\mathbf{u})$$

and therefore can always be calculated from a set of x-ray diffraction data.

Rossmann and Blow (1962) first proposed the rotation function to be a product of 'stationary' (P_0) and 'rotated' (P_1) Patterson functions evaluated within the volume U

$$R(\Omega) = \int_U P_0(x).P_1(\Omega x)\,dx$$

where Ω is the rotation matrix and U is the volume of integration. This function can be expanded as their Fourier series:

$$R(\Omega) = (U/V^3) \sum_{h} (F_{\mathbf{h}}^2 \{ \sum_{p} F_{\mathbf{p}}^2 G_{\mathbf{h},\mathbf{p}} \})$$

where **h** and **p** are the reflection indices corresponding to Pattersons P_0 and P_1 respectively. The interference function G is the Fourier transform of the integration volume. The rotation function can be used in two ways. P_0 and P_1 can be calculated from the same data resulting in a self-rotation function where Ω is a symmetry operation reflecting crystallographic or non-crystallographic symmetry for maximal values of R. If P_0 is calculated from observed data and P_1 is calculated from a model, $R(\Omega)$ is the cross-rotation function where peaks in R may correspond to superimposition of the model onto the unknown structure. Noise is generated in this calculation due to the presence of cross-vectors in P_0 , which may affect the matching of the self-vectors.

Crowther (1972) expanded the Pattersons in the rotation function in terms of spherical harmonics. This allows the use of Fast Fourier Transforms in the calculation, decreasing the execution time, but at the expense

of limiting the volume of integration to being spherical. The radius of integration is inherently linked to the resolution limits by the number of Bessel functions supplied.

Navaza (1987) revised the formulation of Crowther (1972) by using numerical integration instead of Fourier-Bessel expansions in the radial variable. The rotation group Patterson functions are expanded in spherical harmonics but the coefficients of the expansions are evaluated by numerical integration over a spherical surface. Numerical instabilities were also removed by a new algorithm. Consequently, the coupling of the integration radius to resolution limits no longer exists. The rotation peaks are enhanced by the skipping of low angular resolution contributions, but all rotation peaks are explored in a translational search compared to only a single orientation in other molecular replacement packages.

Axel T. Brünger developed molecular replacement procedures within the X-PLOR package. Conventional rotation searches in Patterson space are performed using the real space, rather than reciprocal space, Patterson search method of Huber (1985). The list of rotation function peaks are then filtered via PC-refinement, using a least-squares refinement procedure in a triclinic P1 cell, which should promote 'correct' peaks in relation to 'incorrect' peaks.

6.1.2 Translation Functions

The most widely used method of determining the position of a correctly orientated molecule in the unit cell is the crystallographic residual search:

$$R(\mathbf{t}) = \sum |F_{obs} - kF_{calc}(\mathbf{t})| / \sum F_{obs}$$

where the residual R(**t**) is calculated for all positions (**t**) of the molecule within the asymmetric unit of translation space.

Information on the position of a correctly orientated model relative to a local origin was derived by Crowther and Blow (1967) by evaluating the correlation between a set of calculated Patterson cross vectors and the observed Patterson function of the crystal. Their T function can be expressed as:

$$\mathrm{T}(\mathbf{t_0}) = \int_{\mathrm{v}} \mathrm{P}_{01}(\mathbf{u},\mathbf{t})\mathrm{P}(\mathbf{u})\,d\mathbf{u}$$

Therefore when the calculated cross-Patterson P_{01} fits correctly to the observed Patterson P, $\mathrm{T}(\mathbf{t_0})$ will have a large positive value. The translation function can be achieved by means of a standard Fourier summation by expansion of the Patterson function $\mathrm{P}(\mathbf{u})$ as a Fourier series:

$$\mathrm{T}(\mathbf{t}) = \sum_{\mathrm{h}} |\mathrm{F}_{\mathrm{obs}}(\mathbf{h})^2|\mathbf{F}_{\mathrm{M}}(\mathbf{h})\mathbf{F}^*_{\mathrm{M}}(\mathbf{hA})\exp(-2\pi i\mathbf{h}.\mathbf{t})$$

where $\mathbf{A}$ is the rotational component of the symmetry operation, $\mathbf{F}_M$ the calculated structure factors for the model relative to the local origin and $\mathbf{F}^*_{\mathrm{M}}$ is a complex conjugate of $\mathbf{F}_M$. Intra-molecular vectors can be removed from the observed Patterson decreasing the background noise and yielding the T_1 function:

$$\mathrm{T}_1(\mathbf{t}) = \sum_{\mathrm{h}}(|\mathrm{F}_{\mathrm{obs}}(\mathbf{h})^2| - \sum_{\mathrm{i}=0}^{\mathrm{n}-1}(\mathbf{F}_{\mathrm{M}})(\mathbf{hA_i})^2)$$
$$*\mathbf{F}_{\mathrm{M}}(\mathbf{h})\mathbf{F}^*_{\mathrm{M}}(\mathbf{hA})\exp(-2\pi i\mathbf{h}.\mathbf{t})$$

where $\mathbf{A}_i$ is the rotation matrix of symmetry operator of the i^{th} molecule to the known molecule. T_1 and T functions compare cross-vectors between any pair of symmetry related molecules with the cross vectors in the observed Patterson between all molecules. The T_1 function can be modified to allow matching of all cross vectors between the calculated and observed Pattersons and yielding the T_2 function:

$$T_2(\mathbf{t}) = \int_v (P(\mathbf{u}) - P_M(\mathbf{u})).P_{calc}(\mathbf{u}.\mathbf{t})\,d\mathbf{u}$$

where P_M is the Patterson of the model calculated with respect to the local origin and $P_{calc}(\mathbf{u}.\mathbf{t})$ is the Patterson of the model at a position $\mathbf{t}$ in the unit cell.

6.2 INTRODUCTION

Conceptually the molecular replacement technique is relatively simple if a presumed homologous model is known; determination of the correct orientation of the known model in the cell of the unknown, followed by determination of the translations that best describe the position of the properly orientated known model in the cell of the unknown. But practically many parameters, eg. sequence identity and crystal symmetry, can cause many difficulties in determining the 'correct solution'. Probably the major determining factor is the requirement for a sufficiently similar search model to the undetermined structure (similar of course only with hindsight). The search model usually performs the same function as the unknown and a relatively high sequence identity exists between the two. The unknown in this case is *Tp.acidophilum* citrate synthase with the functionally identical pig heart citrate synthase as the model. There is only 20% sequence identity between the two, although sequence alignments show conservation of the catalytically important residues suggesting similar methods of action and perhaps tertiary fold. As mentioned in Chapter 4, results from circular dichroism spectra suggest similar levels of secondary structure for both enzymes. The RMS differences of Cα atoms between the search model and the refined new structure are usually in the order of 0.5-1.5Å , although higher RMS values (eg.pepsinogen which showed an RMS difference of 1.68Å to the search model (James and Sielecki, 1986)) have been reported. The results presented here are in chronological order, with most detail being reserved for positive features/results. Specific programs used will be detailed in appropriate places.

6.3 RESULTS

6.3.1 Data sets used

Both orthorhombic crystal types were proposed to contain a single dimer within the asymmetric unit, whereas the monoclinic type to contain two dimers. Although identifying the position of one dimer (two monomers) is practically easier than identifying the position of two dimers (four monomers), attempts at solving the monoclinic data set was concentrated on due its superior quality compared to the two orthorhombic data sets (see section 5.5.5). The success of molecular replacement can be dependent on the completeness and quality of data, especially if the search model is known/proposed to deviate from the unknown (eg. low sequence identity) (Schreuder *et al.*, 1992). No results will be presented that are specific to the orthorhombic data sets; only general trends/modifications that are applicable to all data sets.

6.3.2 Search models used

The necessity of having a 'good' search model is an overriding factor in the success (or not) of molecular replacement. A summary of the many slightly differing models used throughout is presented below. There are crystal structures for citrate synthase from only pig heart and chicken heart. The RMS difference between the open forms of the enzyme from the two organisms for main chain atoms is only 0.45Å, with the chicken heart structure having been solved by molecular replacement using the open pig heart structure as the search model (Liao *et al.*, 1991). In the following sections only the use of the native ('open') form/data set of *Tp.acidophilum* citrate synthase will be discussed. The major proposed difference between pig heart and *Tp.acidophilum* citrate synthases as predicted from sequence

alignment studies is the N-terminal extension present in the former. Therefore the N-terminal 53 amino acids of pig heart citrate synthase were deleted from all search models due to their predicted absence in *Tp.acidophilum* citrate synthase. Regions in proteins that may show high flexibility and/or that exhibit no secondary structural characteristics (eg, loop regions) are known to cause problems in finding molecular replacement solutions. The C-terminus of pig heart citrate synthase has little secondary structure and makes only a few interactions to the rest of the molecule , although ARG421 is an active site residue. Therefore an alternative search model was created in which 16 residues were also deleted at the C-terminus. The above two 'deletion' search models were used both as a monomer and a dimer.

The difficulty of finding a 'correct solution' prompted other deletion models to be tested. From crystallographic studies pig heart citrate synthase is known to undergo a large conformation change upon substrate binding (see section 1.1.5) with the small domain rotating 18° with respect to the large domain. Therefore just the large domain of the open form of pig heart citrate synthase was used as a search model, in case the flexibility between the small and large domains was causing problems.

All models were reduced to polyALA/GLY due to the low sequence identity between the model and the unknown. Although this greatly reduces the protein scattering matter present, it has been shown that as little as 1/4 of that present in the asymmetric unit can be successful in obtaining a molecular replacement solution (Fitzgerald, 1988)

6.4 THE SELF ROTATION FUNCTION

Tp.acidophilum citrate synthase exists as a dimer of identical monomers and therefore each monomer must be related to the other monomer by a two-fold operation. The monoclinic data set was thought to contain two dimers within the asymmetric unit and therefore the two dimers must be related to each other by non-crystallographic symmetry. The identification of the NCS operators would aid the success of molecular replacement.

Inspection of the $\kappa = 180°$ should reveal the presence of the expected two-fold dimer axis as well as the crystallographic two-fold axes. Initial attempts at identifying the non-crystallographic symmetry were made using POLARRFN (CCP4), with ϕ and ψ sampled at 5° intervals with κ held at 180°. A range of resolution ranges and the corresponding Patterson cut off radii were employed. The two orthorhombic data sets revealed the expected three crystallographic two-folds and the monoclinic data set the one two-fold, but neither revealed any significant evidence of non-crystallographic two-folds - Fig. 6.1 is the κ=180° section for the csnat134 data set as produced by POLARRFN (CCP4). The rotations enabling the placement of one dimer onto the other dimer, of different orientation, within the asymmetric unit of the monoclinic data set is not constrained to lie on the $\kappa = 180°$ section. Therefore a self-rotation search on csnat134 was performed on κ from 0–180 °, but no obvious significant peak could be identified with confidence. A self-rotation search in Patterson space using X-PLOR on the csnat134 monoclinic data set yielded a peak corresponding to ψ, ϕ and κ angles of 90.0°, 84° and 180° respectively. Very similar values were obtained using CROSUM (MERLOT).

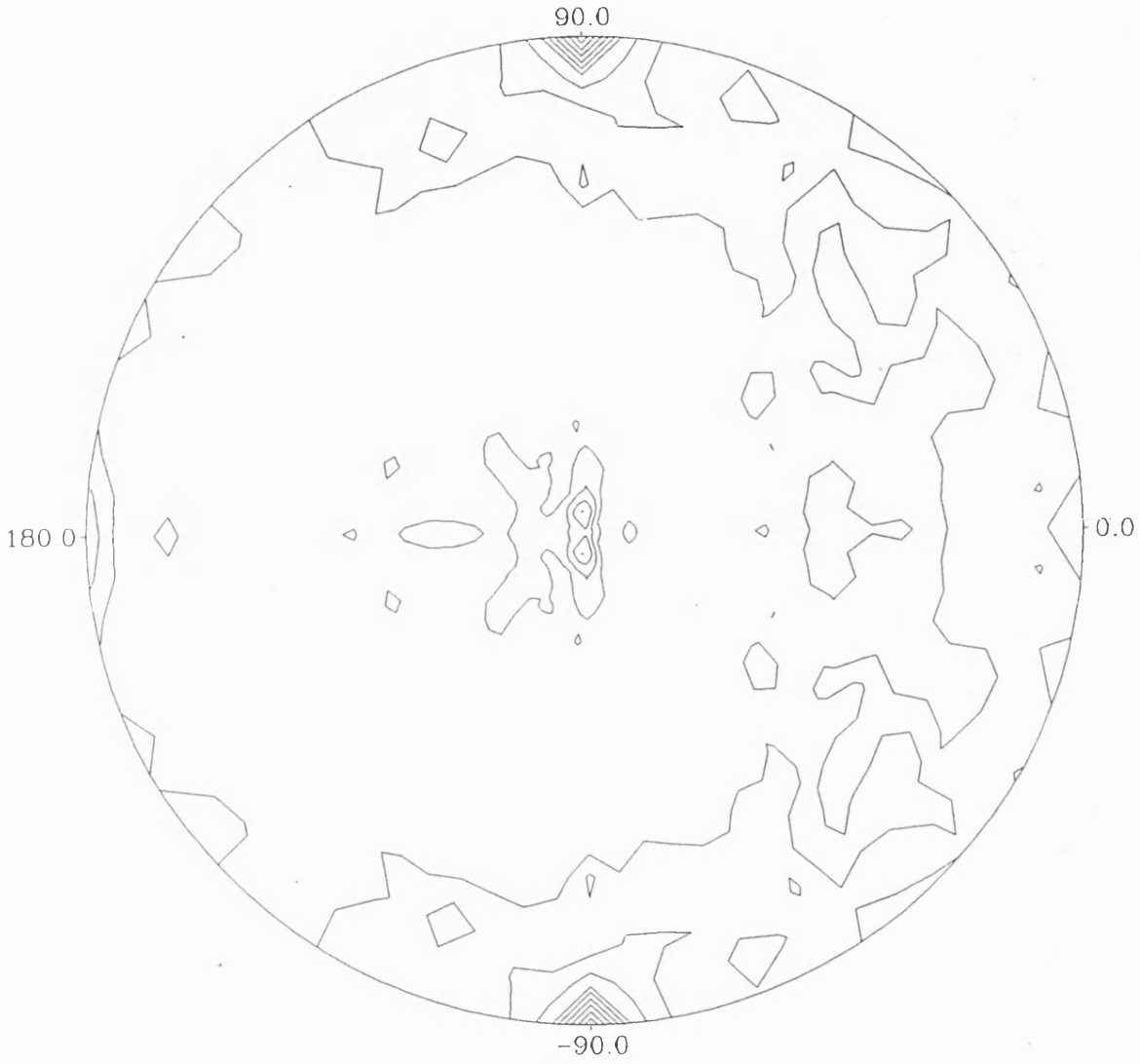


Figure 6.1: A 8-3.0Å self-rotation function on the csnat134 monoclinic data set. Stereographic projection of rotations with $\kappa = 180°$.

6.5 THE CROSS ROTATION AND TRANSLATION FUNCTION

6.5.1 MERLOT

Previously existing molecular replacement techniques were rationalized and integrated into a single package MERLOT (version 2.3) by P.M.D. Fitzgerald (1988).

The rotation function was initially performed using HARMCO and CROSUM (derived from Crowther (1972)) and the positions refined by searching on a finer grid using LATSUM (Lattman and Love 1970). Many resolution ranges were used (10–4Å, 10–6Å and 6-4Å) with the corresponding sphere of integration (MERLOT manual) with all the possible search models but the refined top peak from the rotation function was never greater than 2-3 RMS, which was not very well resolved from the rest of the peaks. The rotation function peaks related by the identified non-crystallographic symmetry were searched using ROTSYM.

Each of the highest peaks from the rotation function was used in the RVAMAP translation procedure, which searches for R-factor correlations, but the resulting list of peaks showed no peak that was significantly higher than the next one on the list.

Many different models were used throughout the use of MERLOT but none of them gave a solution significantly highly than other solutions. Crystal packing of the few possible 'solutions' was tested graphically using FRODO (Jones 1985) ,but all showed very bad packing within the unit cell and were clearly incorrect solutions. Although many different parameters were explored throughout the use of MERLOT no solution could be found to the molecular replacement problem.

6.5.2 X-PLOR

The molecular replacement procedures in X-PLOR (version 2.1) were devised by Axel T. Brünger (1990). As above, a wide range of parameters were investigated in the course of using X-PLOR but only a single example will be detailed below as a test case (the most positive one).

The monoclinic csnat1 data set was used and the N-terminally deleted dimer was used as the search model. Structure factors were calculated for the search model in a P1 cell of dimensions 150x150x150Å with $\alpha=\beta=\gamma=90°$. The rotation function and PC-refinement were performed over 10–4 Å resolution range. PC-refinement was performed using a variety of different 'groups' within the rigid-body minimization procedure. The dimer was treated as a single group, followed by each monomer as a single group and then these were then split into sub-groups consisting of the proposed large and small domains of the monomer. After PC-refinement only a single significant peak was identified - see Fig. 6.2 ; equivalent to the rotation needed for one of the dimers in the asymmetric unit. The translation position for this dimer was then identified, with the Y coordinate of the first molecule fixing the origin along the unique b axis, although the top peak was only about 1σ above the mean. The position of the second molecule was identified relative to the first one using the first position which had been rotated by the non-crystallographic symmetry operator (see section 6.4), with the whole of the asymmetric unit explored in the translation search . The second position, again, was only about 1σ above the mean.

This 'solution' was checked initially for crystal packing with symmetry related molecules using the graphics package FRODO - see Fig. 6.3. A few surface loops showed interpenetration but was judged to be sufficiently 'correct' for an initial simulated-annealing refinement protocol (see

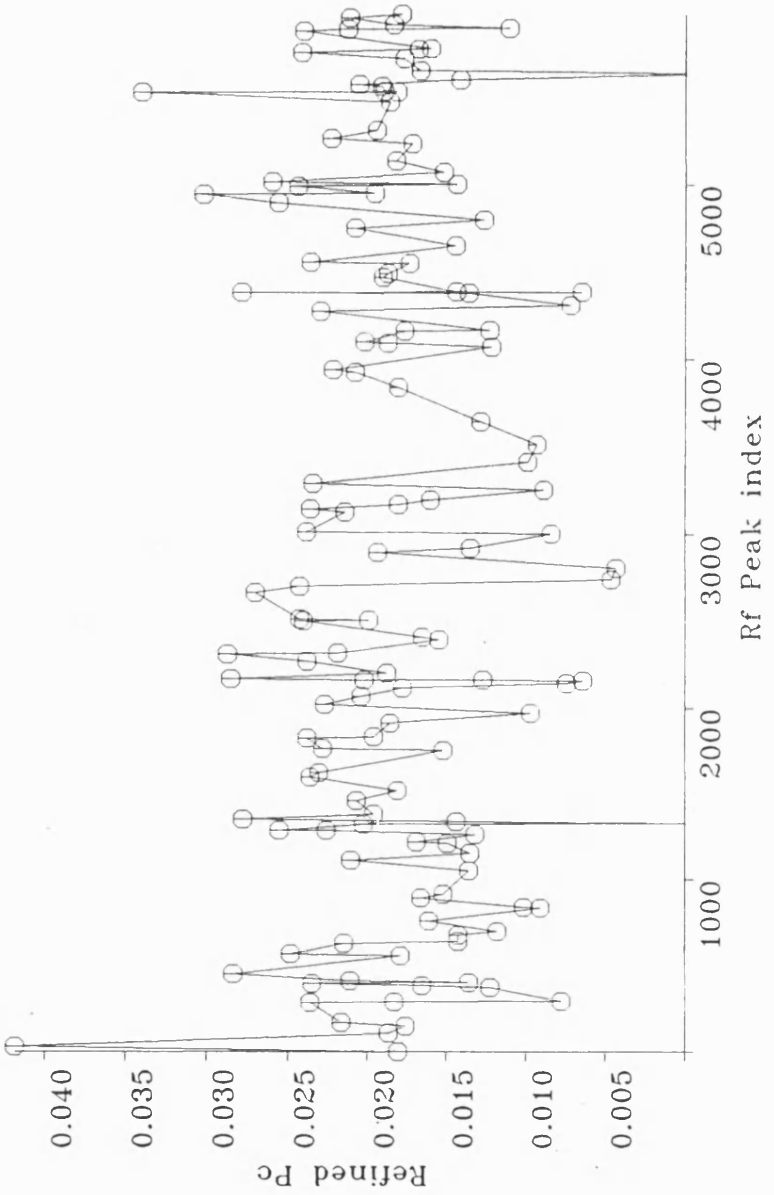


Figure 6.2: Cross-rotation peaks for the csnat134 data set, produced from X-PLOR after PC-refinement.

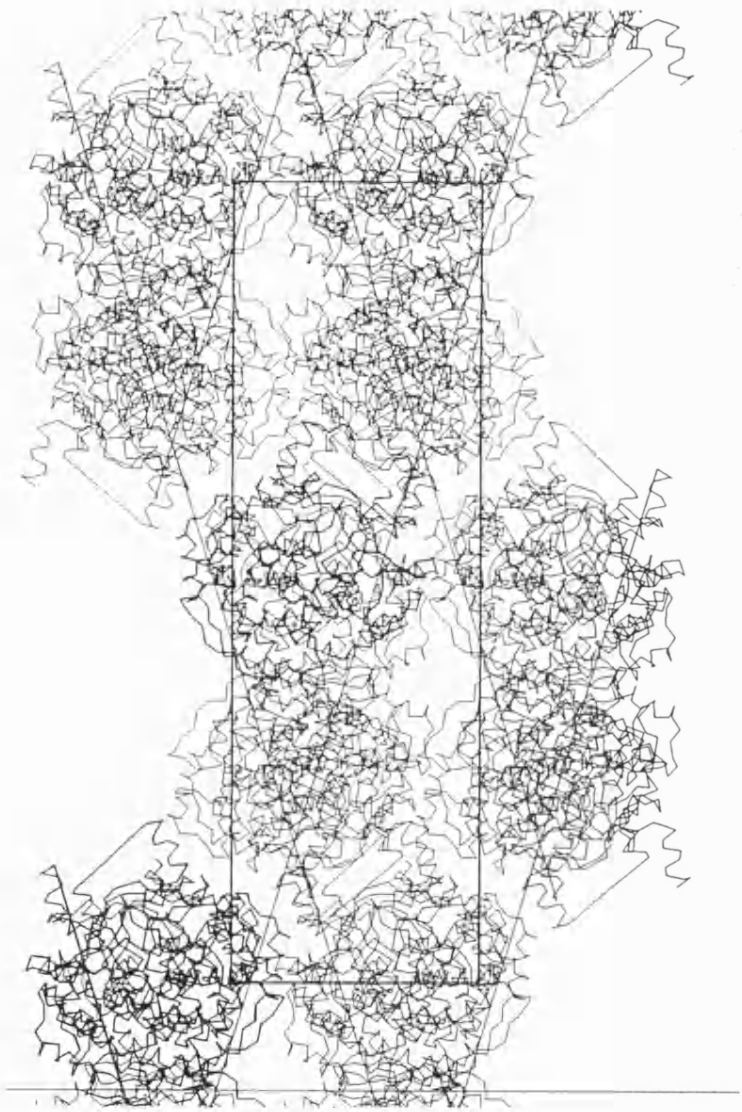


Figure 6.3: Molecular packing of the search model within the unit cell of type2 crystals, as predicted from the incorrect molecular replacement solution from X-PLOR. The view is down the c axis. Produced by FRODO (Jones, 1985)

Chapter 8) to 2.8Å resolution limit to proceed. An initial $2F_o$-F_c map was generated using the phases from the molecular replacement solution and areas of side-chain density was apparent in the interpretable map. A R-factor of 28.8% was reached but without imposing NCS-restraints. Upon imposition of these parameters the R-factor rose to 42.5%. Inspection of the individual monomers revealed an RMS difference between Cα atoms for supposedly identical monomers was around 2.5Å , if no NCS-restraints were imposed. Therefore this solution was deemed to be incorrect and a return to molecular replacement was made.

6.5.3 AMORE

AMORE is a package for molecular replacement devised by Jorge Navaza (1992). The new strategy developed is the fast automatic translational exploration of a large number of rotation function solutions instead of a single rotation function peak, which has been post-rotationally refined as in X-PLOR. The individual sequence of the stages within AMORE are summarised in Fig. 6.4. The first stage, SORTING, is the packing and sorting in a P1 asymmetric unit of the hkl Fobs. Next, in TABLING, the search model is placed in a small box, the coordinates rotated so that the principal axes of inertia are parallel to the box axes and translated so that the centre of gravity is at the origin; ensuring an optimal placement of the model within a minimal box. TABLING also produces arrays of Fourier coefficients of the model densities corresponding to the search model. Once the input data have been transformed into a suitable representation by the above two stages, ROTING is able to interpolate the structure factors for each orientation of the model in a suitable cell from the Fourier coefficients produced in TABLING. Spherical harmonics for both the crystal and model are also calculated in ROTING which are used for the Rotation function

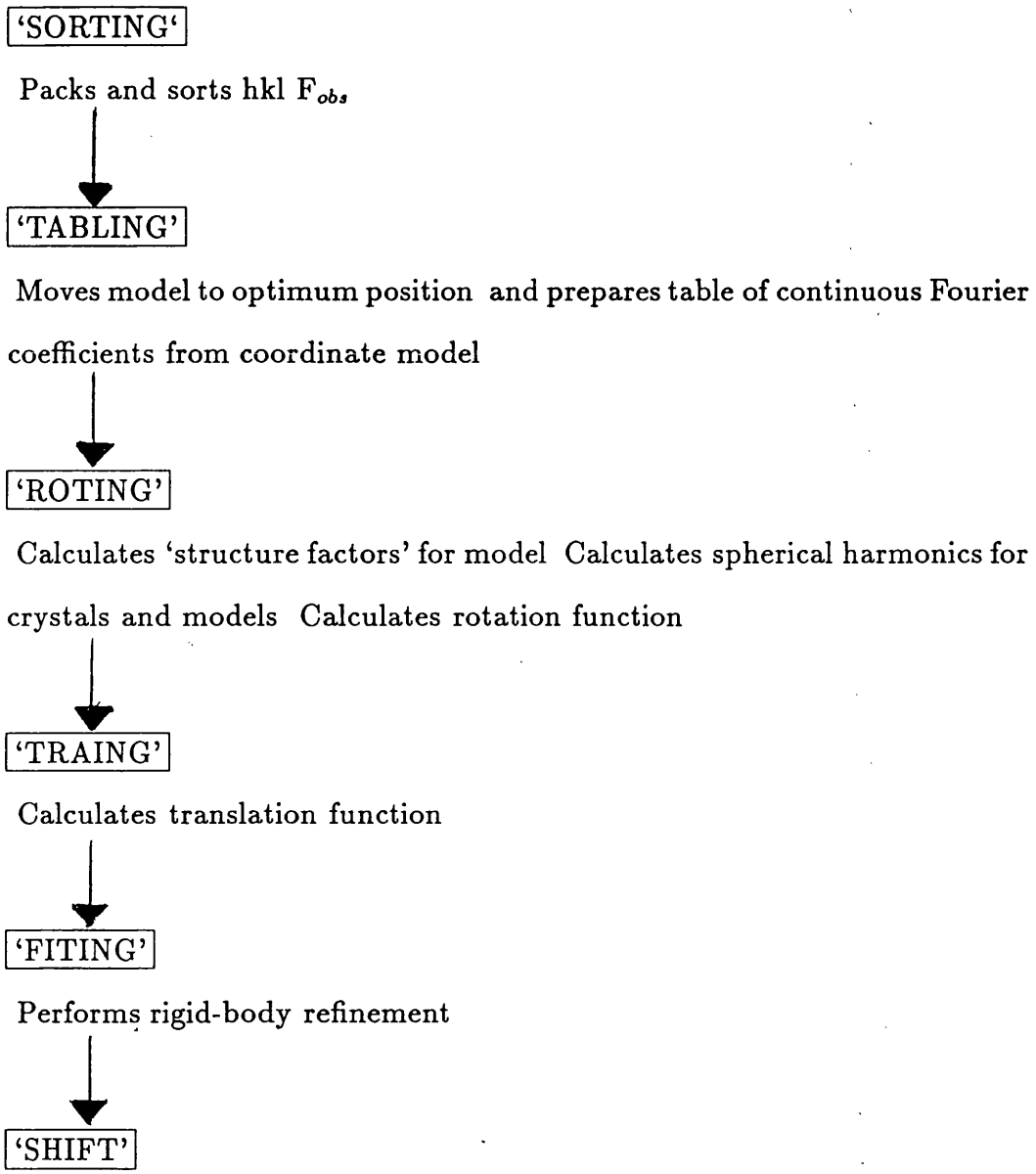


Figure 6.4: Flow chart summarising the individual programs with AMORE, the molecular replacement package by Navaza (1992).

calculation. If a suitable rotation function peak is identified, the translation part of the molecular replacement calculation can be carried out using TRAING, using the Crowther and Blow overlap function (1967) method. The rotation and translation peaks can be optimized using FITING; a rigid-body refinement by the method of Castellano *et al.* (1992), an improved version of the Huber and Schneider method (1985). Since the search model coordinates are rotated/translated into an optimal position in the search box (in TABLING), the final stage in AMORE is the calculation, from the refined parameters produced by FITING, of the final rotation and translation parameters to be applied to the search model.

The model chosen for use in the AMORE package was the pig heart citrate synthase dimer which had been both N- and C- terminally deleted, which was used in conjunction with the monoclinic data set csnat134. All the different rotation functions were carried out in a P1 cell with dimensions 150x150x150Å with $\alpha=\beta=\gamma=90°$. The resolution limits used were 20Å– 4Å. Higher resolution data were not included due to size dependent computing limits and the structural differences likely due to 20% sequence identity between unknown and search model.

The first variable investigated was the sphere of integration radius, over which the Patterson maps were to be calculated. Navaza suggests that this should not be greater than the maximal distance from the center of mass, 48.98Å as calculated by TABLING. R_{min} was constant as zero with R_{max} (the Patterson cut off radii) being tried at 20Å, 25Å and 30Å. The radius of integration must be optimised so as to reduce the contribution of intermolecular vectors to the Patterson calculation. The subsequent rotation function solutions are listed in Table 6.1 (only the top nine peaks are listed for simplicity). The top peak was the same for both 25Å and 30Å

α	β	γ	Correlation	Integration sphere
				0–25
243.28	79.96	5.11	15.3	
21.50	122.42	143.42	13.2	
133.24	125.48	320.06	12.4	
120.04	63.73	31.46	12.1	
26.50	46.19	148.50	11.9	
34.00	48.29	143.50	11.9	
302.05	58.91	33.46	11.5	
65.18	86.12	182.44	11.3	
331.61	90.12	153.52	11.2	
				0–30
242.91	79.99	5.31	12.9	
148.50	133.50	326.00	11.6	
24.01	124.16	323.94	11.3	
113.76	94.68	2.29	11.3	
298.98	59.87	213.07	10.3	
116.78	63.21	34.31	10.3	
336.18	123.50	332.76	9.6	
18.74	122.43	331.89	9.6	
				0–20
152.75	2.40	107.48	19.3	
620.00	0.00	0.00	18.8	
0.00	180.00	-100.00	18.8	
0.00	180.00	-280.00	18.8	
40.20	56.42	324.63	17.8	
298.50	26.00	359.00	17.6	
25.94	122.38	322.85	17.2	
205.14	66.57	238.03	17.1	

Table 6.1: List of cross-rotation function peaks on the csnat134 data set with the N- and C- terminally deleted pig heart citrate synthase as the search model, as calculated by ROTING in AMORE. Only the top nine peaks as listed for simplicity. The Patterson radius of integration is the variant between the three different searches. All searches were performed using 10-4Å data.

radii, but list of rotations used for subsequent translation searches were calculated with a 25Å cut off radius due the higher correlation present for the top peaks.

The Crowther and Blow translation method (1967) was used throughout. The peaks from the rotation function were initially used to find the relative position of one of the dimers in the asymmetric unit, with b being the unique axis. The same resolution limits were imposed on the translation function as were on the rotation function. The sampling interval for the translation map (SHANN) was initially set to 2 but failed to give high correlation values for the resulting peaks. But on reduction of this interval to 1, higher correlation values were obtained - see Table 6.2 for the list of peaks. This reduction causes the 'coordinate map' to be sampled on a coarser grid. Peak number 8 gave both the highest correlation value and the lowest R-factor and therefore was taken as the position of the first dimer within the asymmetric unit.

Peak No.8 was fixed and the translation function run again on the same set of rotation peaks with the position to be determined relative to the position of the first dimer. Table 6.3 shows the output from the second translation function. The solution for molecule 2 with the highest correlation value and lowest R-factor is equivalent to the top peak from the rotation function, and was therefore taken as the position of the second dimer within the asymmetric unit.

With a possible solution for the position of the 2 dimers a rigid body refinement was carried out using the method of Castellano *et al.* (1992) to refine the rotation and translation solutions, allowing monitoring of the correlation and R-factor. The final refined positions are stated in Table 6.4. The R-factor dropped and the correlation coefficient rose upon the

Peak No.	α	β	γ	a	b	c	Correlation	R-factor
1	243.28	79.96	5.11	0.06667	0.00000	0.08333	15.0	56.7
2	21.50	122.42	143.42	0.36667	0.00000	0.08333	14.4	56.5
3	133.24	125.48	320.06	0.25000	0.00000	0.42708	14.2	57.0
4	120.04	63.73	31.46	0.20000	0.00000	0.26042	13.9	57.0
5	26.50	46.19	148.50	0.18333	0.00000	0.22917	13.7	56.7
6	34.00	48.29	143.50	0.25000	0.00000	0.07292	14.3	56.3
7	302.05	58.91	33.46	0.01667	0.00000	0.16667	14.6	56.7
8	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
9	331.61	90.12	153.52	0.40000	0.00000	0.44792	12.1	57.0
10	356.88	116.27	215.42	0.26667	0.00000	0.39583	13.4	57.0
11	315.97	176.47	212.09	0.41667	0.00000	0.42708	13.4	56.3
12	30.36	78.64	301.40	0.18333	0.00000	0.10417	12.8	57.0
13	203.86	61.81	60.14	0.11667	0.00000	0.48958	13.2	56.5
14	172.71	114.81	212.57	0.30000	0.00000	0.43750	11.8	57.7
15	273.54	123.60	246.83	0.10000	0.00000	0.41667	14.2	56.7
16	143.50	154.95	318.50	0.31667	0.00000	0.30208	12.6	56.9
17	1.56	88.87	87.61	0.23333	0.00000	0.48958	12.8	56.8
18	94.51	121.20	159.89	0.23333	0.00000	0.01042	13.9	56.8
19	91.71	151.29	151.00	0.30000	0.00000	0.37500	11.5	57.5
20	330.16	147.74	322.98	0.11667	0.00000	0.33333	12.8	56.9
21	0.00	180.00	256.90	0.41667	0.00000	0.42708	12.6	56.9

Table 6.2: List of translation function peaks after one of the cross-rotation function peaks had been fixed, as calculated by TRAING in AMORE. The search was limited to x and z, with 10-4Å data being used for the calculation.

	α	β	γ	a	b	c	Correlation	R-factor
1	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	243.28	79.96	5.11	0.56144	0.24010	0.96634	20.6	55.2
2	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	21.50	122.42	143.42	0.74806	0.86890	0.00000	18.0	56.1
3	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	133.24	125.48	320.06	0.50776	0.58067	0.69708	17.6	56.2
4	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	120.04	63.73	31.46	0.19385	0.88805	0.98323	17.5	55.7
5	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	26.50	46.19	148.50	0.84304	0.47523	0.05443	17.8	55.9
6	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	34.00	48.29	143.50	0.20696	0.01156	0.36849	17.9	56.4
7	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	302.05	58.91	33.46	0.72491	0.09664	0.30314	17.6	56.3
8	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	331.61	90.12	153.52	0.35039	0.90430	0.88745	16.2	56.1
9	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	356.88	116.27	215.42	0.27649	0.58658	0.09854	16.9	56.3
10	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	315.97	176.47	212.09	0.03818	0.61822	0.41235	16.8	56.0
11	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	30.36	78.64	301.40	0.80915	0.85547	0.52208	17.2	56.1
12	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	203.86	61.81	60.14	0.32248	0.46960	0.48482	16.9	56.3
13	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	172.71	114.81	212.57	0.06169	0.07934	0.70032	16.4	56.7
14	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	273.54	123.60	246.83	0.40719	0.05250	0.24001	17.3	56.2
15	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	143.50	154.95	318.50	0.36012	0.39364	0.51918	16.5	56.0
16	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	1.56	88.87	87.61	0.71703	0.41197	0.49321	17.0	56.1
17	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	94.51	121.20	159.89	0.42883	0.13632	0.63384	17.9	56.0
18	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	91.71	151.29	151.00	0.49020	0.37559	0.30406	16.8	55.9
19	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	330.16	147.74	322.98	0.38214	0.60837	0.71638	16.5	56.3
20	65.18	86.12	182.44	0.15000	0.00000	0.28125	16.9	56.0
	0.00	180.00	256.90	0.33681	0.51401	0.10626	16.9	55.9

Table 6.3: List of translation peaks for the position of the second dimer in the asymmetric unit with respect to the fixed position for the first dimer, as calculated by TRAING in AMORE. 10-4Å data were used in the calculation.

α	β	γ	a	b	c	Correlation	R-factor
68.11	86.71	181.34	0.16795	0.00059	0.28875	50.1	50.8
248.05	77.75	5.46	0.54676	0.23826	0.96568	50.1	50.8

Table 6.4: Orientations and positions of the two dimers after rigid-body refinement of the initial molecular replacement solutions produced by AMORE.

α	β	γ	a	b	c	Correlation	R-factor
164.36	27.81	307.94	5.85	-0.13	24.83	50.1	50.8
184.05	153.65	163.55	19.16	41.63	83.03	50.1	50.8

Table 6.5: Orientations and positions of the two dimers within the asymmetric unit of crystal type2.

refinement suggesting a possible 'solution'. The final R-factor was still over 50% but since only a polyALA search model was used the lack of scattering matter would be an accountable factor.

In TABLING the search model is rotated so that its principal axes of inertia are parallel to the box axes and then translated so that the centre of gravity is at the origin. Therefore the above solutions were converted so that they applied to the original search model. The rotation and translation functions that were applied to the pig heart citrate synthase dimer are stated in Table 6.5. The molecular replacement solution for the position of the two dimers was checked initially for packing with symmetry related molecules within the unit cell using the graphics package FRODO - see Fig. 6.5. Although minor clashes were present these were confined to a few surface loop regions, with the majority of the dimer showing no bad packing contacts with either the other dimer in the asymmetric unit or with symmetry related dimers. Therefore this 'solution' was deemed to be a good candidate for further refinement procedures (see Chapter 8) to check whether there was truth in the 'solution'.

6.6 DISCUSSION

Molecular replacement can be a very rapid and effective tool in the elucidation of crystal structures, but in this instance it was more of a case of the latter than the former. Although the sequence identity between pig heart and *Tp.acidophilum* citrate synthase is only 20% there was data that indicated similar levels of secondary structure between the two enzymes (see Chapter 4). The success of molecular replacement is dependent on the level of structural similarity between the unknown and the search model. Therefore, the reduction of the search models to a polyALA/GLY chain due to the low sequence identity made the search model less than ideal.

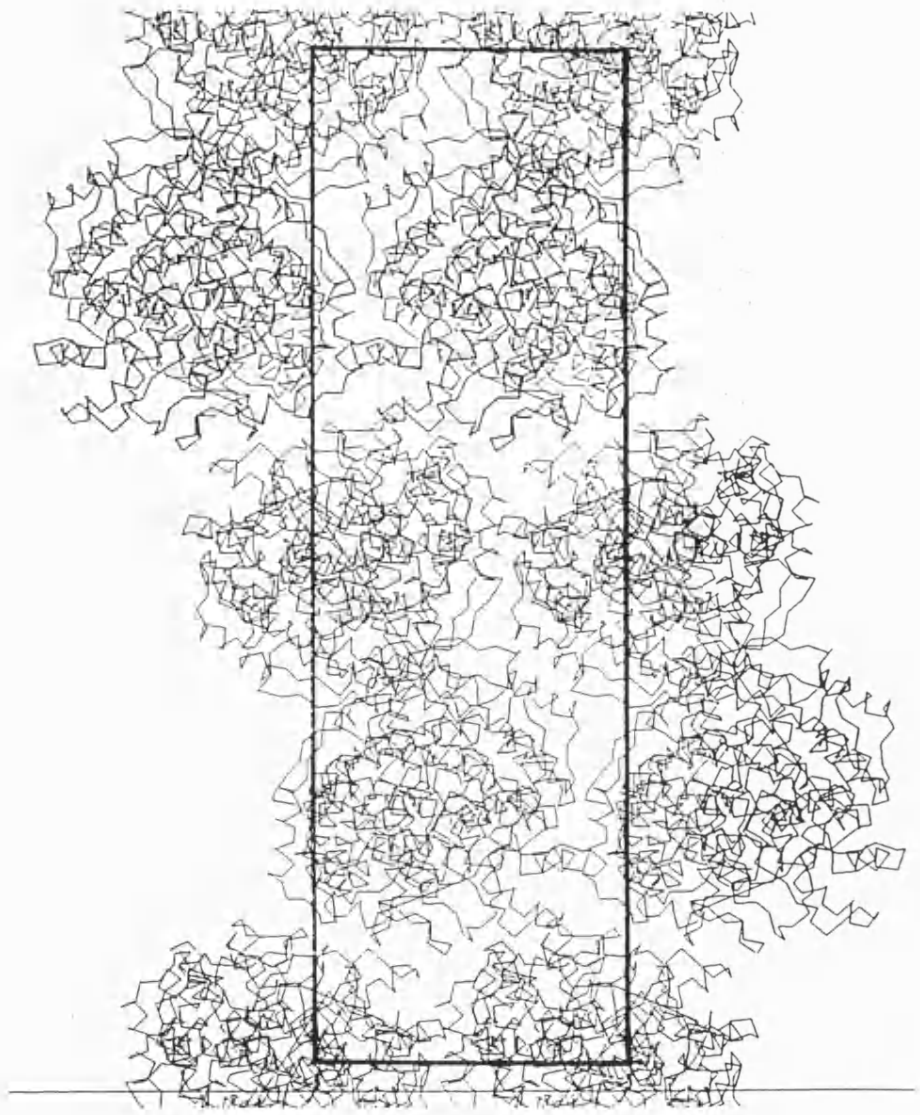


Figure 6.5: Molecular packing of the search model within the unit cell of type2 crystals, as predicted from the molecular replacement solution from AMORE. The view is down the c axis. Produced by FRODO (Jones, 1985).

A wide range of molecular replacement packages were used to the full, exploiting most possible parameters available but only AMORE gave the correct solution. It is an intriguing question as to why AMORE worked and the other packages didn't. One feature of AMORE is the ability to explore all rotation function peaks in the same translation function search, although the top peak was in fact equivalent to the orientation of one of the dimers. This function was also useful since the NCS operators could not be identified with great confidence. The orientation of one of the dimers was equivalent to the eighth peak of the original rotation function.

The data sets for crystal types 1 and 3 proved to be unsuccessful in yielding a molecular replacement solution. This was probably due to the low quality of the data coupled with the poor (ie. low sequence identity) search model. The data for crystal type 2 was of far superior quality and coupled with the use of AMORE was probably the significant factor in gaining a 'correct' solution.

Although no bad contacts in the molecular packing of the solution from X-PLOR were present and a low R-factor was reached upon initial refinement, the solution was deemed to be incorrect since good agreement between observed and calculated data was only seen at the expense of NCS. Brady and Jian-sheng (1992) previously reported that simulated annealing refinement procedures must be used with caution when testing molecular replacement solutions, especially with respect to the imposition of NCS-restraints.

The lack of confidence in the detection of NCS may be explained. The peak at $\phi = 90°$, ie. parallel to the unique axis b, is the crystallographic 2-fold, for the monoclinic data set. and therefore detection of NCS along b is impossible. Wang and Janin (1993) reported that if a dimeric protein

crystallizing in space groups like $P2_1$ or $P2_12_12_1$ and it lacks strict twofold symmetry, that with a high frequency crystal packing will orientate the local twofold relative to the crystal screw axis. Therefore if the local twofold axis of *Tp.acidophilum* citrate synthase is orientated parallel to b, detection of the former will be impossible. Although this does not explain the case of the monoclinic data set where the self rotation should identify the rotations to orientate one dimer onto the second dimer. Wigley (1992) had a similar problem in identifying the relationship between two tetramers within the asymmetric unit with the self rotation function. He interpretated this to be due to the two tetramers to be in very similar orientations.

The refined positions of the molecular replacement solutions allows identification of the relationship between one monomer and the other as well as between the two dimers within the asymmetric unit. The rotation, in spherical polar angles, between one monomer and the second monomer of the dimer is ψ=157, ϕ = 171 and κ=180, which corresponds to a rotation of 180° around an axis with direction cosines 0.37, 0.92 and 0.0. The rotation between the two dimers in the asymmetric unit is ψ=69, ϕ=177 and κ=194, which corresponds to a rotation angle of 165° about an axis with direction cosines 0.9, -0.35 and 0.0. Therefore upon inspection of the κ=165° section a peak should be present which corresponds to the NCS between the two dimers in the asymmetric unit. A peak only 3σ above the origin peak was present on this section and therefore initial confident identification of this peak was not possible. Interestingly the relationship between a monomer of one dimer to the two-fold related monomer of the other dimer in the asymmetric unit corresponds to a rotation of ψ=94, ϕ=90 and κ=180, ie. the monomers in the two dimers are also related by a two-fold.

Chapter 7 : The Search for heavy atom derivatives

7. 1 INTRODUCTION

In the absence of a molecular replacement solution, a screen of heavy atom soaks was instigated in the search for isomorphous heavy atom derivatives. From the binding of a heavy atom(s) to a protein new scattering centres are introduced into the protein causing a change in the intensity of reflections yielding information concerning the unknown phases. Not unlike the planning of crystallization trials, the choice of which heavy atom to use may be seen as a mixture of judgement with hindsight and instinct. Unfortunately many heavy atoms do not bind specifically at a small number of sites within the specified protein, or show no binding at all. Many also cause a large perturbation of the protein, which if used in crystal soaking experiments will disrupt the crystal lattice leading to either non-isomorphous data (ie. the cell dimensions being altered and/or rotation of the molecule in the same unit cell) or the preformed crystal cracking upon soaking.

Derivative data sets need to be isomorphous with the native data set. If any of the cell dimensions have changed by more than 1% of the native cell dimensions a large amount of 'noise' is created in the subsequent analysis due to the large changes in the intensity of the reflections. A 0.5% change in all of the cell dimensions results in a mean change of 15% in reflection intensity at 3Å resolution (Blundell and Johnson, 1976).

The theory behind the heavy atom isomorphous replacement method for the calculation of phases will not be discussed in this chapter due to no phase information being gained from the results presented below.

7.2 RESULTS

Initial attempts to co-crystallize protein and heavy atom solution proved unsuccessful, therefore crystal soaking experiments were employed where a previously grown native crystal was soaked in a heavy atom solution of known concentration for a specified time. The 'soaked' crystals were then mounted in a quartz capillary tube and diffraction data collected. The chosen heavy atom was dissolved in the PEG solution used for the crystallizations. Many of the heavy atoms chosen as potential candidates did not dissolve eg.$AgNO_3$, $PbAc_2$ and $CdCl_2$. Fortunately many of the heavy atoms did dissolve and were used as candidates for soaking experiments.

The analysis was carried out using the CCP4 suite of programs. Native and derivative data sets were combined using MTZUTILS, which was followed by anisotropic and local scaling (LOCAL) to bring them onto the same relative scale. An isomorphous difference Patterson map was generated (FFT) and the Harker sections investigated for peaks. The Harker sections were plotted using PLUTO. VECSUM was used for the semi-automatic interpretation of the Patterson. A first potential site ($>2\sigma$) on the Harker section was chosen, which fixed the hand and origin. This site was fed back into VECSUM, and subsequent possible sites were added until no further significant peaks were obtained. The positions of the peaks corresponding to the potential heavy atom sites were refined using VECREF. The peaks were extensively statistically examined for useful phase information using MLPHARE in terms of figures of merit and phasing power, for example.

Table 7.1 summarises the results from the soaking experiments. No useful isomorphous heavy atom derivative was found for either crystal type. The Patterson maps never yielded a peak above 3σ and the sites that were

Table 7.1: Summary of heavy atom soaking experiments.

Heavy atom	(mM)	Soak time (hr)	Effect	Data collected	Crystal type	Isomorphous
Uranyl acetate	5	12	Crystal cracked	NO	—	—
K_2PtCl_6	5	24	Crystal disordered	NO	—	—
$Th(NO_3)_4$	2	22	Crystal disordered	NO	—	—
Na_3IrCl_6	4	120	Crystal disordered	NO	—	—
$NaAuCl_4$	5	0.2	Crystal cracked	NO	—	—
UF	6	25	—	YES	2	NO
$La(CH_3COO)_3$	2	20	—	YES	2	YES
K_2PtCl_4	5	24	—	YES	2	YES
K_2PtCl_4	5	48	—	YES	3	YES
KI_3	0.4	24	—	YES	3	NO

present proved to be low in phasing power.

Due to no heavy atom derivatives being found detailed information on results from only a single heavy atom, K_2PtCl_4, will be presented which will be used as a model example for why statistically the specific heavy atom was not a useful derivative.

Single crystals (type2) were soaked in 5mM K_2PtCl_4 for 24 hr. The cell dimensions of the derivative showed a 0.3% change in a, 0.3% change in b, 0.3% change in c and 0.4% change in β, a derivative sufficiently isomorphous for subsequent analysis. The data set was 82% complete to 6Åand 45% complete to 4.5Å. Analysis was carried out to both the above resolutions but data will only be presented on data to 6Å. Figure 7.1 is the isomorphous difference Patterson map on the single Harker section (v=1/2). No significant peaks were present, suggesting lack of specific binding of the heavy atom but further analysis was carried out to gain statistical information on the possible derivative.

4 peaks were identified from VECSUM and the position of these sites in the unit cell were refined using VECREF. The residual values for the 4 sites are very high (0.7–0.8) and the correlation quite low (0.5–0.6), indicative of a low quality derivative. From the refined heavy atom positions, the single isomorphous replacement phase information was calculated using mlphare. The mean figure of merit for phased acentric measurements is 0.1688 and for phased centric measurements 0.3339 (on a 0–1 scale, reflecting the accuracy of the phase angle). A fom of 1 is the ideal value. Thus the value for this derivative was low, again indicative of a low quality derivative. The phasing power of the derivative was 0.5, which again is a low value. A value of 1 or over is indicative of a derivative which may give good information concerning the phases. No anomalous data was present.

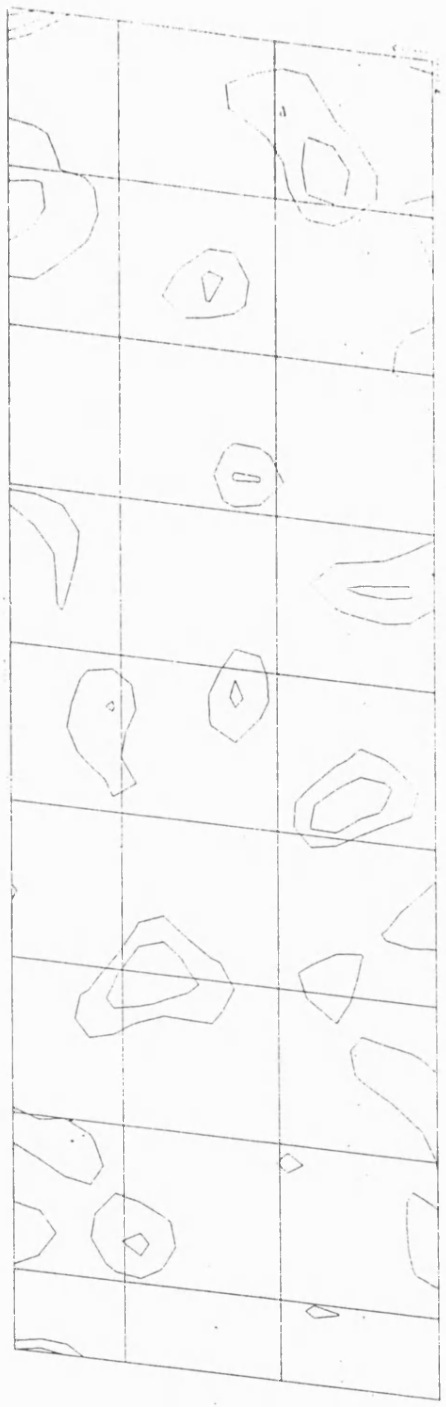


Figure 7.1: Isomorphous difference Patterson map on the Harker section v=1/2, calculated from K_2PtCl_4 (5mM, 24hr) soaked type2 crystal and csnat134 data sets. The map was contoured in levels of σ, starting at 1σ.

7.2 DISCUSSION

The search for isomorphous heavy atom derivatives can be a long and arduous task and in this instance proved to be unsuccessful for both the crystal types (2 and 3) tried. K_2PtCl_4 was soaked into both the crystal forms but neither type resulted in a derivative suggesting, in this instance that the crystal packing was not a factor in heavy atom binding but rather the nature of the heavy atom itself, in relation to the protein. In both crystal types the perturbation of the crystal caused by soaking in a heavy atom solution was a major problem resulting in non-isomorphism or crystal cracking/disordering.

The phasing statistics for K_2PtCl_4 as a derivative are poor. The equivalent phasing statistics for the other heavy atoms analysed have similar values. Therefore, unfortunately, no useful heavy atom derivatives for *Tp.acidophilum* citrate synthase were found, although the success of molecular replacement eventually eradicated their need.

Why no useful heavy atom derivatives were found is partly due to the intrinsic nature of the protein. Some heavy atoms are known to bind to specific amino acids in proteins, which of course must be accessible to the heavy atom. An amino acid commonly cited as a binding site for heavy atoms is cysteine, due to the presence of the sulphur atom which is able to react rapidly and irreversibly with mercuric ions; the positive charge of the mercury reacting with the negative charge of the sulphydryl group via nucleophilic attack. Unfortunately *Tp.acidophilum* citrate synthase contains no cysteine residues (although no mercury salts would actually dissolve in the PEG crystallization solution). This lack has been a problem in the search for derivatives of other proteins, one which was only overcome by the mutation of a cysteine residue into the protein and the subsequent use

of the mutant protein for determination of the crystal structure via heavy atom (ie. mercuric) derivatives (Nagai *et al.*, 1991).

Useful heavy atom derivatives may have been found upon investigation of different heavy atoms, more extensive investigation of concentrations and soak times, but this is more of a matter for conjecture than of necessity.

Chapter 8 : Refinement of *Thermoplasma Acidophilum* citrate synthase structure

8.1 REFINEMENT THEORY

One method of crystallographic refinement of macromolecules can be seen as a non-linear minimization of a hybrid energy function (Jack and Levitt, 1978). Stereochemical restraints are included in the energy minimization calculations to improve the effective number of observations since macromolecular diffraction data are usually not in the atomic resolution range. Convergence of the energy term by gradient descent minimization will only occur if the structure being refined is close to being optimal since only local energy minima will be found. The recent use of molecular dynamic simulations into refinement procedures enables larger energy barriers to be overcome and a wider range of conformations to be explored in the search for a global energy minimum. Other methods of refinement (eg PROLSQ and TNT) involve a series of stereochemical restraints not expressed energetically, but as a series of distance and angle restraints.

The progression of the refinement is usually followed by calculation of the agreement between the calculated structure factor amplitudes and the observed ones, known as the R-factor where

$$R = \sum_{h} F_o(\mathbf{h}) - F_c(\mathbf{h}) / \sum_{h} F_o(\mathbf{h})$$

Therefore as the agreement between the observed amplitudes and the calculated amplitudes increases (ie. as your model increasingly fits to the observed electron density), the value of the R-factor decreases. Known fully refined protein structures have an R-factor only between 15-20%, due to practically unavoidable errors in the collection and processing of the x-ray data and , more importantly, the inability to describe the full contents

of the unit cell which typically contains 50% solvent.

8.2 THE X-PLOR ENERGY FUNCTION

The effectiveness of the refinement procedures carried out by X-PLOR is based on definition of the X-PLOR energy function, which can be grouped into two different classes:

$$E_{TOTAL} = E_{EMPIRICAL} + E_{EFFECTIVE}$$

$E_{EMPIRICAL}$ describes the energy of the molecule(s) through an empirical energy function derived from stereochemical restraints on the molecule. $E_{EFFECTIVE}$ comprises restraining energy terms that use experimental information or other information.

8.2.1 The Empirical Energy Function

The empirical energy function is made up of the following stereochemical terms:

$E_{EMPIRICAL} = E_{BONDS} + E_{ANGLES} + E_{DIHEDRALS} + E_{IMPROPERS} + E_{NONBONDED}$

The first four terms are conformational energy terms.

$E_{BOND} = \sum_{bonds} k(r - r_0)^2$

describes the covalent bond energy, where r is the observed bond distance, r_0 is a 'standard' bond distance, derived from small molecule structures, and k is a force constant derived from spectroscopic studies of bond stretching.

$E_{ANGL} = \sum_{angles} k(\theta - \theta_0)^2$

describes the bond angle energy, where θ is the observed bond angle, θ_0 is a 'standard' bond angle derived from small molecule structures and k is a force constant derived from spectroscopic studies.

$$E_{IMPR} = E_{DIHE} = \sum_{dihedrals,impropers} \begin{cases} k(1 + \cos(n\phi + \delta)) \text{ if } n > 0 \\ k(\phi - \delta)^2 \text{ if } n = 0 \end{cases}$$

describes energy terms involving dihedral angles, chirality or planarity. n is the periodicity of the dihedral, δ is the ideal dihedral angle and ϕ is the observed dihedral angle.

The nonbonded interactions are described by:

$$E_{NONBONDED} = E_{VDW} + E_{ELEC} + E^P_{VDW} + E^P_{ELEC}$$

where the third and fourth terms are equivalent functions to the first and second terms but describe interactions between symmetry related molecules. The nonbonded term can be switched to a purely repulsive function to reduce numerical instabilities in the refinement of initial structures which may contain bad contacts. The hydrogen-bond energy is contained within the electrostatic and van der Waals terms.

The electrostatic term is given by

$$E_{ELEC} = \sum_{i<j} Q_i Q_j \frac{C}{\epsilon_o R}(1 - \frac{R^2}{R^2_{off}})^2 + e_{14} \sum_{(i,j)\in\{1-4\}} Q_i Q_j \frac{C}{\epsilon_o R}(1 - \frac{R^2}{R^2_{off}})^2$$

and the Van der Waals term by

$$E_{VDW} = \sum_{i<j}(A/R^{12} - B/R^6) * SWITCH(R, R_{on}, R_{off})$$

$$+ \sum_{(i,j)\in\{1-4\}} (A/R^{12} - B/R^6) * SWITCH(R, R_{on}, R_{off})$$

8.2.2 The Effective Energy Function

The effective energy function consists of terms from experimental observations:

$$E_{EFFECTIVE} = E_{XRAY} + E_{NCS}$$

The XRAY term is split into distinct contributions from the amplitude and phase:

$$E_{XRAY} = E^A_{XRAY} + E^P_{XRAY}$$

where

$$E^{A}_{XRAY} = W_A/N_A \sum_{\mathbf{h}} \mathbf{W}_{\mathbf{h}}[F_{obs}(\mathbf{h}) - kF_{calc}(\mathbf{h})]^2$$

and

$$E^{P}_{XRAY} = W_P/N_P \sum_{\mathbf{h}} \mathbf{W}_{\mathbf{h}}\mathbf{S}\{(\phi_{obs}\mathbf{h} - \phi_{calc}\mathbf{h}), \arccos(m(\mathbf{h}))\}$$

W_A and W_P are the determined ideal weights on the structure factor amplitude and phases relative to the geometric terms, with $W_{\mathbf{h}}$ the weight on the individual observations. E^{P}_{XRAY} is only employed if phase restraints are used. k is a scale factor to ensure that E^{A}_{XRAY} is a minimum and N_A is a normalization factor to make the weight independent of the resolution range.

Non-crystallographic symmetry (NCS) restraints are introduced into the X-PLOR effective energy function by restraining NCS-related atoms to their average position by means of an effective harmonic energy term, equivalent to the method used by Hendrickson and Konnert (1980) in their refinement program. A least-squares superposition of these atoms onto a reference set (taken as the first set defined) is computed, and the average x,y,z ($\bar{x}$) for each atom is taken. Each atom is then restrained according to the energy:

$$E_{NCS} = w(x - \bar{x})^2$$

with w used to weight the restraint. NCS related B-factors are similarly restrained to their average value :

$$(b - \bar{b})^2/sigb^2$$

where $\bar{x}$ and $\bar{b}$ are the mean values from the NCS-related atoms.

8.3 MOLECULAR DYNAMICS

Molecular dynamics consists of simultaneously solving Newton's equations of motion for all atoms under the influence of a potential derived from empirical stereochemical terms and observational restraints. In terms

of the X-PLOR energy function the equation takes the form of:

$$m_i \frac{d^2 x_i}{dt^2}(t) = -\nabla_{x_i} E_{TOTAL}$$

A genuine 'free' molecule dynamics simulation is not achieved due to the incorporation of observational restraints into the dynamics equation. Extra energy is incorporated into the system by the assignment of velocities to atoms allowing local energy minima to be traversed. Kinetic energy is introduced into the system via the release of stereochemical restraints and the increased agreement with the experimental data. The velocities of the atoms give the current temperature of the system by Boltzmann's Law. The maintenance of the temperature of the system can be controlled in one of three ways. Firstly, the velocities can be periodically rescaled in an uniform fashion by

$$v_i^{new} = v_i^{old}(T/T_{curr})^{1/2}$$

for all atoms i where T is the target temperature and T_{curr} is the current temperature. This method of temperature control is used in the 'check' procedure for standard molecular dynamic simulations. Secondly, the influence of a heat bath can be incorporated into the classical equations of motion by Langevin dynamics, which applies random forces to the friction coefficient. The third option is the weak temperature–coupling method, which is related to Langevin dynamics but instead applies a global scale factor to the friction coefficient which is proportional to the ratio of T_{curr} to T. In the 'slowcooling' simulated annealing protocols used in X-PLOR, the T-coupling method is used in preference to Langevin dynamics due to faster convergence rates seen in the former since the friction coefficient becomes zero if T_{curr} agrees with T , which does not slow the atomic motions as much as when using Langevin dynamics. Upon using the T-coupling method for temperature maintenance the X-PLOR dynamics equation becomes

$$m_i(\delta^2 r_i/\delta t^2) = -\nabla_{x_i} E_{TOTAL} - m_i \gamma_i v_i (1 - (T/T_{CURR}))$$

where γ is the friction coefficient.

8.4 RIGID BODY REFINEMENT

The X-PLOR method of rigid body refinement minimizes the six rotational and translational degrees of freedom of each specified group of atoms, that is, the groups of atoms are treated as rigid bodies. The complete energy function E_{TOTAL} is used.

8.5 REFINEMENT STRATEGY

8.5.1 Rigid body refinement

The first step after the determination of the possible molecular replacement solution was the optimization of the orientation and position of the model within the asymmetric unit by rigid body refinement using the program X-PLOR (Brünger *et al.*, 1987). Since the orientation and position of 4 identical monomers were identified, non-crystallographic restraints were employed to ensure that each monomer was treated as identical, with each monomer being an equivalence set within a group.

8.5.2 Refinement by Simulated Annealing

The protocol used for a simulated annealing refinement (program X-PLOR; Brünger *et al.*, 1987) is shown in Fig. 8.1. The first stage of the protocol is to explicitly build in the polar hydrogens and to create a Protein Structure File ('psf'), a molecular topology file containing information about atom names, angle terms and explicit hydrogen-bonding terms for example. Subsequent refinement procedures require a weighting between E_{XREF} and $E_{EMPIRICAL}$ obtained using the 'check' procedure. The value

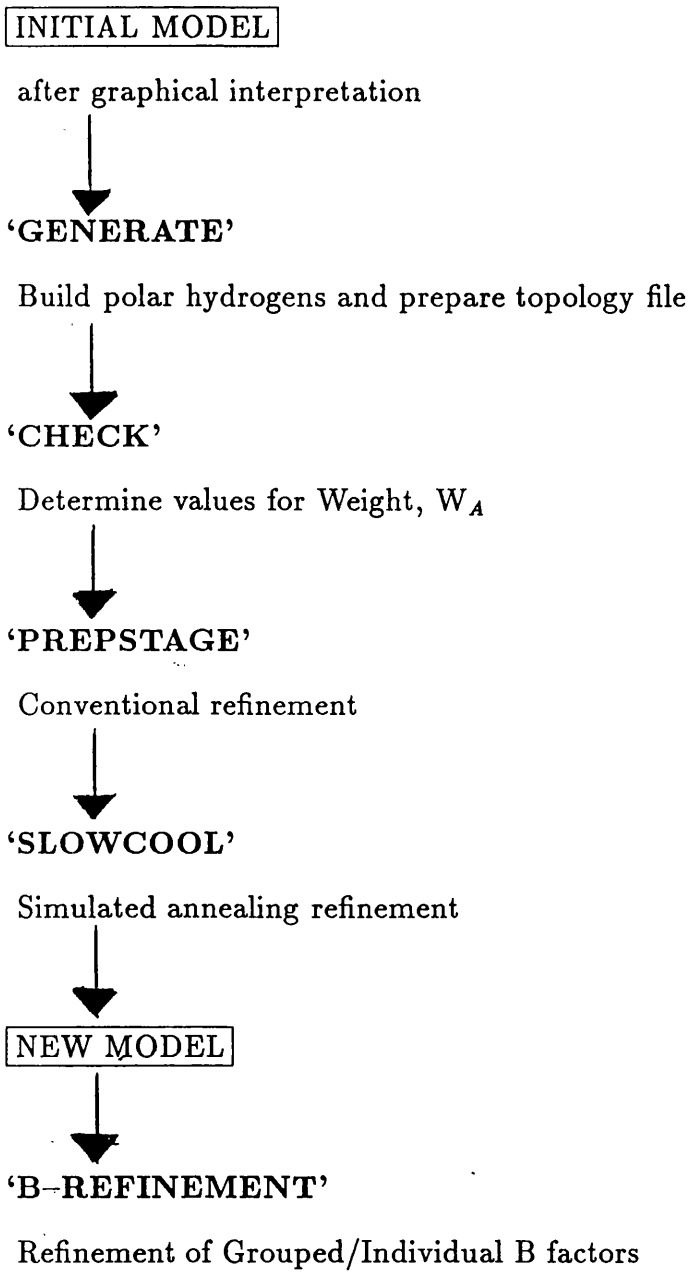


Figure 8.1: Flow diagram summarising the individual stages in the simulated annealing refinement protocol employed.

is obtained at the start of each refinement cycle by 40 steps of Powell minimization and a 0.1ps molecular dynamics calculation (300K) with E_{XREF} excluded. The gradients of E_{XREF} and $E_{EMPIRICAL}$ are then compared in order to give the ideal value for W_A.

NCS-restraints were employed in refinement procedure due to the presence of four monomers within the asymmetric unit. When the majority of the sequence had been fitted and refinement was at an advanced stage, the NCS-restraints were removed.

Before simulated annealing refinement using the coordinates of all free atoms as variables, energy minimization must be carried out to relieve strain or bad contacts, using the 'prepstage' procedure. To ensure convergence in minimization of the initial structure the 'repel' nonbonded energy function was used for 40 cycles which uses a harmonic repulsive term instead of the vdw term, followed by a further 120 cycles of Powell minimization with the vdw term switched on in the energy minimization calculation. The purely repulsive minimization cycles were not necessary as the refinement progressed due to fewer bad contacts being present within the molecule.

Simulated annealing refinement is carried out in the 'slowcool' stage. The initial temperature was set to 3000K (or 2000K to reduce computational time) and reduced at 0.5fs intervals to 300K in order to find a global energy minimum. At each temperature interval 50 steps of dynamics were carried out. After the molecular dynamic refinement 120 cycles of Powell minimization were performed to alleviate any strain or bad contacts that may have arisen from the 'slowcooling'.

The model used in AMORE was pig heart citrate synthase which had been reduced to polyALA/GLY. Therefore the *Tp.acidophilum* citrate synthase sequence had to be fitted manually using the graphics package O

(versions 5.7, 5.8 and 5.9) (Jones *et al.*, 1991). Sequence alignment of pig and *Tp.acidophilum* citrate synthases (BESTFIT, GCG) (Fig. 8.2) were used to identify equivalent regions of sequence and these gave a starting position for initial analysis of the contiguous regions of electron density. The graphical mutations were performed on a single monomer and the other three monomers converted at the end of the manual mutations by least-squares transformations using O.

After each round of graphical interpretation a complete round of simulated annealing refinement was undertaken, followed by creation of a new $2F_O$-F_C map. All electron density maps were calculated with a 2σ cutoff using the program X-PLOR (Brünger) and converted to O format using esmappage on the Evans and Sutherland ESV-10.

8.6 REFINEMENT PROGRESSION

Rigid body refinement was carried out on the csnat134 data set over a 10–4 Å resolution range. Each monomer, treated as a single rigid body, was initially subjected to 20 cycles of energy minimization which was then followed by 50 cycles of minimization treating each monomer as two distinct rigid bodies – equivalent to the large and small domains of pig heart citrate synthase. The R-factor at this stage of refinement was 53.4%.

The refinement statistics from the cycles of simulated annealing refinement are stated in Tables 8.1 and 8.2 and will be discussed in more detail below. Initial rounds of refinement were performed using csnat134 data set in the resolution range 8-2.8 Å using a 2σ cut off. The initial $2F_O$-F_C electron density map was of variable quality, since only a fraction of the known scattering matter was present in the starting model. Although many regions of density were not contiguous, regions were present in which the presumed backbone density was uninterrupted with areas of side chain

```
            1                                                  50
         Pig  ASSTNLKDIL ADLIPKEQAR IKTFRQQHGN TVVGQITVDM MYGGMRGMKG
Thermoplasma  .......... .......... .....PETEE ISKGLEDVNI KWTRL.....
            1                                                  50
         Pig  ASSTNLKDIL ADLIPKEQAR IKTFRQQHGN TVVGQITVDM MYGGMRGMKG
Thermoplasma  .......... .......... .......... ....PETEE. ISKGLEDVNI

            51                                                 100
         Pig  LVYETSVLDP DEGIRFRGYS IPECQKMLPK AKGGEEPLPE GLFWLLVTGQ
Thermoplasma  ....TTIDGN KGILRYGGYS VEDI.....I ASGAQD...E EIQYLFLYGN
            51                                                 100
         Pig  LVYETSVLDP DEG IRFRGY SIPECQKMLP KAKGGEEPLP EGLFWLLVTG
Thermoplasma  KWTRLTTIDG NKGILRYGGY SVEDIIAS.. G......AQD EEIQYLFLYG

           101                                                 150
         Pig  IPTEEQVSWL SKEWAKRAAL PSHVVTMLDN FPTNLHPMSQ LSAAITAL.N
Thermoplasma  LPTEQELRKY KETVQKGYKI PDFVINAIRQ LPRESDAVAM QMAAVAAMAA
           101                                                 150
         Pig  QIPTEEQVSW LSKEWAKRAA LPSHVVTMLD NFPTNLHPMS QLSAAITALN
Thermoplasma  NLPTEQELRK YKETVQKGYK IPDFVINAIR QLPRESDAVA MQMAAVAAMA

           151                                                 200
         Pig  SESNFARAYA EGIHRTKYWE LIYEDCMDLI AKLPCVAAKI YRNLYREGSS
Thermoplasma  SETKFKW... ...NKDTDRD V....AAEMI GRMSAITVNV YRHIMNMPAE
           151                                                 200
         Pig  SESNFARAYA EGIHRTKYWE LIYEDCMDLI AKLPCVAAKI YRNLYREGSS
Thermoplasma  ASETKF.K.. ......WNKD TDRDVAAEMI GRMSAITVNV YRHIMNMPAE

           201                                                 250
         Pig  IGAIDSKLDW SHNFTNMLGY TDAQFTELMR LYLTIHSDHE GGNVSAHTSH
Thermoplasma  LPKPSDSYAE SFLNAAFGRK ATKEEIDAMN TALILYTDHE VP.ASTTAGL
           201                                                 250
         Pig  IGAIDSKLDW SHNFTNMLGY TDAQFTELMR LYLTIHSDHE GGNVSAHTSH
Thermoplasma  LPKPSDSYAE SFLNAAFGRK ATKEEIDAMN TALILYTDHE V.PASTTAGL

           251                                                 300
         Pig  LVGSALSDPY LSFAAAMNGL AGPLHGLANQ EVLVWLTQLQ KEVGKDVSDE
Thermoplasma  VAVSTLSDMY SGITAALAAL KGPLHGGAAE AA...IAQFD EIKDPAMVEK
           251                                                 300
         Pig  LVGSALSDPY LSFAAAMNGL AGPLHGLANQ EVLVWLTQLQ KEVGKDVSDE
Thermoplasma  VAVSTLSDMY SGITAALAAL KGPLHGGAAE AAIAQFDEIK ...DP.AM..

           301                                                 350
         Pig  KLRDYIWNTL NSGRVVPGYG HAVLRKTDPR YTCQREFALK HLPHDPMFKL
Thermoplasma  WFND...NII NGKKRLMGFG HRVYKTYDPR AKIFKGIAEK LSSKKPEVHK
            301                                                 350
         Pig  KLRD.YIWNT LNSG.RVVPG YGHAVLRKTD PRYTCQREFA .LKH.LPHD.
Thermoplasma  .V.EKWFNDN IINGKKRLMG FGHRVYKTYD PRAKIFKGIA EKLSSK..KP

           351                                                 400
         Pig  VAQLYKIVPN VLLEQGKAKN PWPNVDAHSG VLLQYYGMTE M.NYYTVLFG
Thermoplasma  VYEIATKLED FGIKAFGSKG IYPNTDYFSG IVYMSIGFPL RNNIYTALFA
           351                                                 400
         Pig  PMFK.LVAQ. LYKI.VPNVL LEQGKAKNPW PNVDAHSGVL LQYYGMTEMN
Thermoplasma  EVHKVYEIAT KLEDFGIKAF G....SKGIY PNTDYFSGIV YMSIGFPLRN

            401                                                443
         Pig  VSRALGVLAQ LIWSRALGFP LERPKSMSTD G....LIKLV DSK
Thermoplasma  LSRVTGWQAH FIEYVEEQQR LIRPRAVYVG PAERKYVPIA ERK
           401                                                 450
         Pig  .YYTVLFGVS RALGVLAQLI WSRALGFPLE RPKSMST..D GLIKLVDSK
Thermoplasma  NIYTALFALS RVTGWQAHFI EYVEEQQRLI RPRAVYVGPA ..ERKYVPIAERK
```

Figure 8.2: Sequence (top) and structural alignment (bottom) of *Tp.acidophilum* citrate synthase and pig heart citrate synthase, produced by BESTFIT, GCG (Devereux *et al.*, 1984) and SHP (Stuart, 1979) respectively.

Table 8.1: Statistics of the progression of the refinement of *Tp.acidophilum* citrate synthase.

Refinement cycle	No. of atoms	No. of reflections	Resolution	R-factor	NCS-Restraints	B-factor refinement
	8740	34433	2.8 Å	39.9%	YES	NO
1	9648	34433	2.8 Å	37.8%	YES	NO
2	10252	34433	2.8 Å	36.3%	YES	NO
3	10688	34433	2.8 Å	34.8%	YES	NO
4	11432	34433	2.8 Å	34.0%	YES	NO
5	12168	34433	2.8 Å	32.6%	YES	NO
6	11984	34433	2.8 Å	31.1%	YES	NO
7	12196	34433	2.8 Å	30.0%	YES	NO
8	12324	34433	2.8 Å	29.7%	YES	NO
9	12916	34433	2.8 Å	23.3%	NO	B-GROUP
10	13420	34433	2.8 Å	22.4%	NO	B-GROUP
11	14176	45875	2.5 Å	23.5%	NO	B-INDIVIDUAL
12	14776	45875	2.5 Å	23.5%	NO	B-INDIVIDUAL

Refinement cycle	Residues mutated
	—
1	109-148, 401-418
2	90-108, 376-400
3	225-237, 245-275
4	370-375, 312-338
5	149-224
6	60-89
7	276-288, 238-244, 54-59
8	337-369
9	289–311
10	36-53
11	419-437
12	—

Table 8.2: Residues (pig heart numbering) mutated into the structure of *Tp.acidophilum* citrate synthase on each round of refinement.

density apparent. The pig heart citrate synthase (reduced to polyALA, ie. the search model) was superimposed onto the map, via the rotational and translational parameters identified in AMORE, and backbone regions identified that showed high degrees of fitting to the contiguous density.

The rounds of residue fitting and subsequent refinement progressed gradually with, on average, a 2% reduction in the R-factor after each round. On average 40-60 residues were mutated at each round of graphical interpretation due to low confidence in determining the matching of the search model to certain stretches of density. Therefore residue replacement erred on the side of caution, especially during initial refinement cycles. The first regions of side chain density to be identified were equivalent to residues 390-410 and 110-140 in pig heart citrate synthase. The former containing an active site residue and the latter involved in subunit-subunit contacts. After these residues were mutated to the equivalent ones in *Tp.acidophilum* citrate synthase, as judged by sequence alignments, a round of refinement was carried out. The electron density map after this showed improved areas of side chain density in regions adjacent to the previously mutated regions. In this way the mutation/refinement cycles progressed in a bootstrapping fashion, until all 384 residues had been inserted into the model. Both the N- and C- termini showed the poorest regions of electron density, with areas at the C-terminus showing non-contiguous backbone density and therefore were the last residues to be added to the model. Both these regions involved the addition of residues to the model since the terminally deleted model had been used to solve the molecular replacement calculation.

NCS-restraints were dropped on refinement cycle 9 when the majority of the *Tp.acidophilum* residues had been inserted into the model. Since the two monomers of a single dimer are not related by a strict crystallographic

two fold, they would not be expected to have identical conformations. After the first round of refinement without NCS-restraints imposed, the RMS difference between individual monomers was in the order of 0.6-0.7 Å over Cα atoms, which considering statistical errors in the data was in the right range for the restraints to be dropped fully. On the same round of refinement, a cycle of isotropic grouped B-factor refinement was employed. After this round of refinement, with no NCS-restraints and B-group refinement, the R-factor reduced dramatically from 29.7% to 23.3%, which finally reduced to 22.4% upon addition of more side chain atoms.

After all the residues had been fitted, the refinement was extended to a resolution limit of 2.5Å with individual B-factor refinement being possible , due to the use of more data in the calculations. The R-factor rose slightly to 23.5% due to use of the higher resolution data being included in the calculation. Once this refinement had been completed a Ramachandran plot was produced on the model allowing identification of regions in the structure which were conformationally strained. Manual rebuilding of these areas, mainly the N- and C- termini and surface loop regions, was carried out and another round of refinement was performed. No reduction was seen in the R-factor but a small improvement in the overall geometry of the structure and a reduction in the number of 'disallowed' residues was observed. With hindsight refinement should possibly not have been extended to 2.5Å since the data to that resolution is weaker and only about 45% complete. The latter causes the individual B-factor refinement employed at this stage to be taken with some caution since the ratio of observations to number of parameters is less than 1. The refinement stopped at this stage due to time constraints.

Fig 8.3 shows a typical section of electron density. There is a high degree

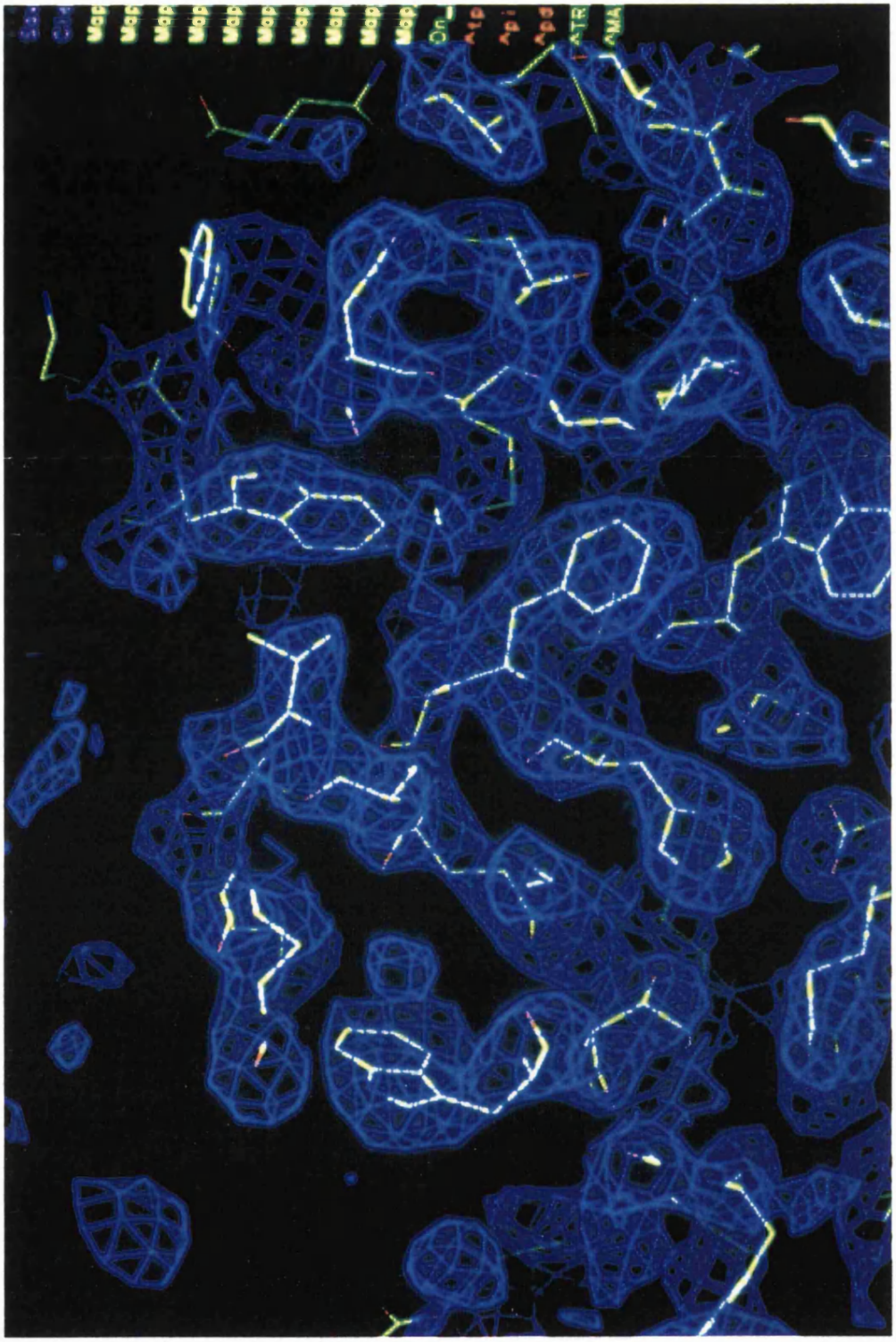


Figure 8.3: $2F_O$-F_C electron density map for a typical region of the partially refined structure of *Tp.acidophilum* citrate synthase. The map was contoured at 1σ and created in O (Jones, 1991).

of fitting between the observed density and the model, indicative of a good quality model. Both map and model presented were produced after the final round of refinement. In order to evaluate the fitting of the structure to the density on a residue by residue basis, RSFIT in O was carried out on the partially refined structure. This program calculates the real space fit of a structure to an electron density map and outputs the correlation coefficient of each residue to the electron density it occupies. Fig. 8.4 displays the fitting of the partially refined *Tp.acidophilum* citrate synthase to a $2F_o$–F_c map calculated after the final round of refinement. It clearly displays that the majority of the structure has high correlation coefficients, but there are certain regions that show a low degree of fitting and therefore these regions deserve some more attention in the rounds of manual rebuilding and refinement.

The quality of the model was initially assessed using PROCHECK (version2.1). The Ramachandran plot for the partially refined *Tp.acidophilum* citrate synthase is shown in Fig. 8.5. There are quite a large number of residues in the generous or disallowed regions of the plot, indicating that the structure is still only partially refined. The majority of the residues in disallowed regions are either in subunit contacts or in extended surface loops, for example, and therefore may be forced into a conformationally strained position, in the case of the surface loop residues by possible crystal contacts. Further work is necessary to ensure as many residues as possible are in the favoured/allowed regions of the Ramachandran plot.

Fig. 8.6 displays a plot of average main chain B-factor values for *Tp.acidophilum* citrate synthase. There are no regions within the protein that show very high flexibility, as judged by B-factor values. The N- and C- termini show high values relative to the rest of the protein, which may

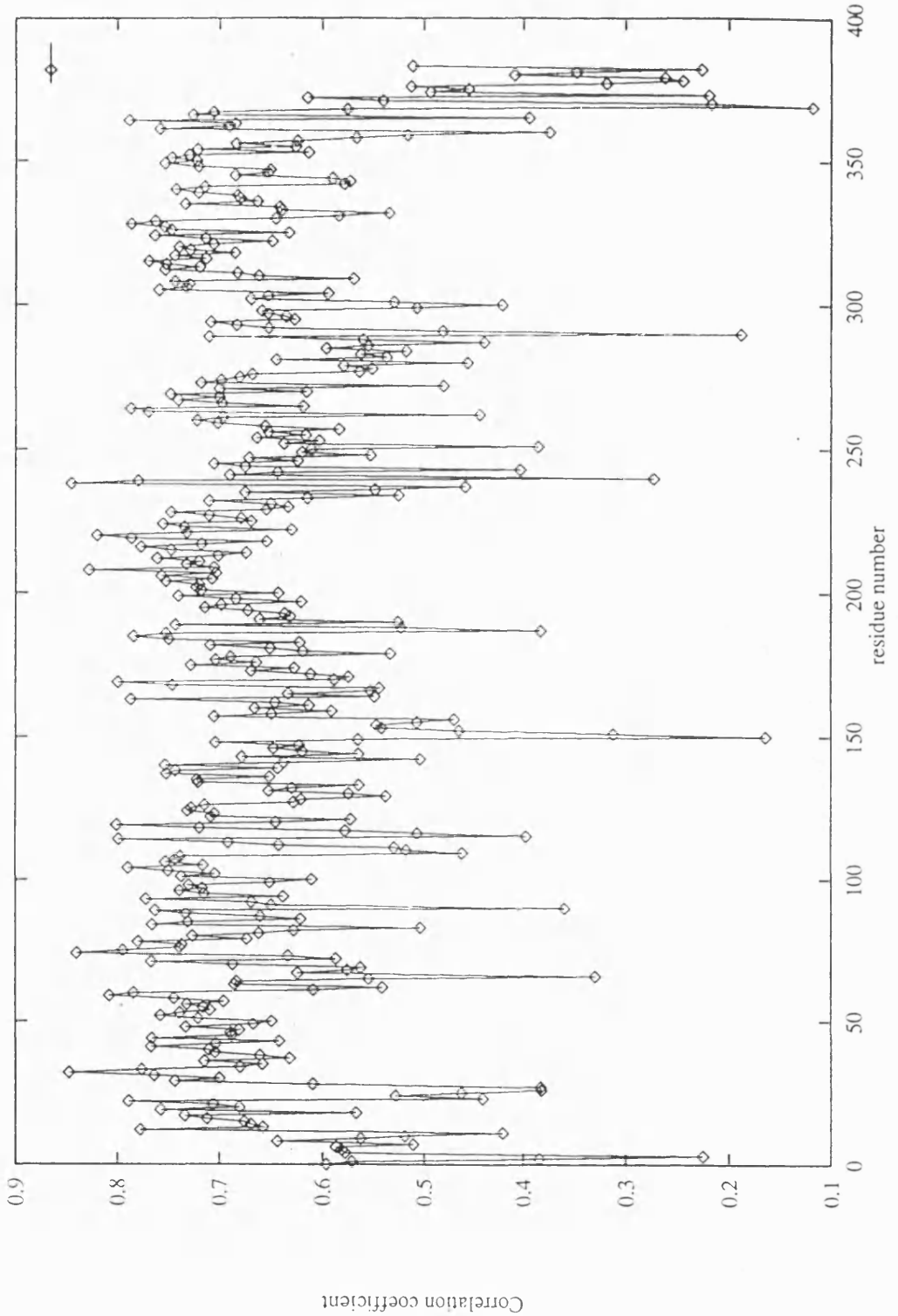


Figure 8.4: Plot of correlation coefficient of the fitting of *Tp.acidophilum* citrate synthase to the final $2F_o$–F_c map. Created using RSFIT in O (Jones *et al.*, 1991).

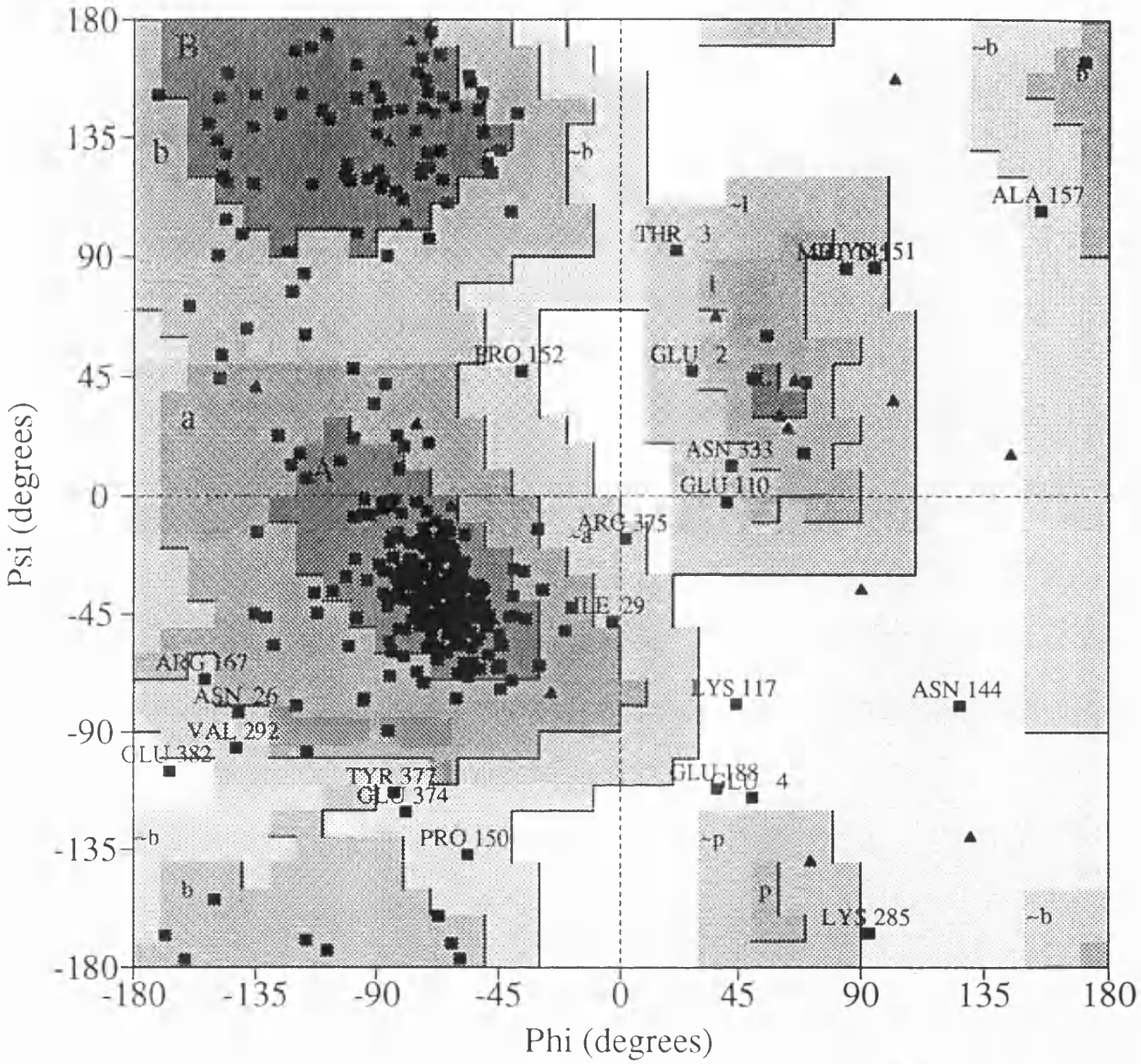


Figure 8.5: Ramachandran plot for a monomer of the partially refined structure of *Tp.acidophilum* citrate synthase. Glycine residues are shown as triangles.

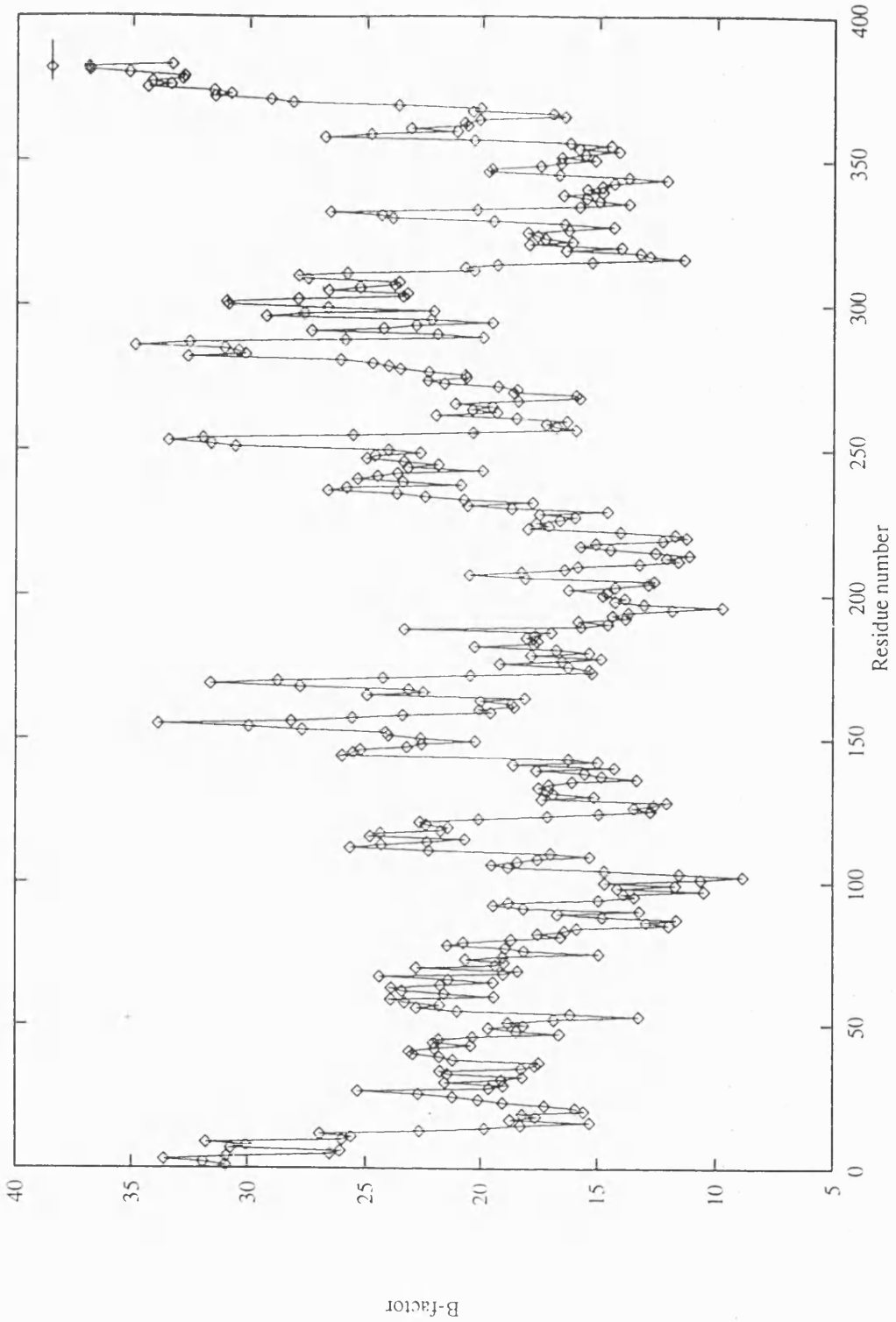


Figure 8.6: Plot of average main chain B-factors (Å^2) for a monomer of the crystal structure of *Tp.acidophilum* citrate synthase.

be due to the unrefined nature of these areas owing to the poor quality of the electron density or that these regions are flexible in the native protein resulting in the poor quality of the electron density. Figs. 8.7 and 8.8 display expected ranges for specific main chain and side chain parameters, all of which lie within the expected values. Many of the parameters are worse than average, specifically the peptide bond planarity. Again this is probably due to the unrefined nature of the protein. Therefore further refinement is necessary with the existing data set, although improved and extended native data collection remains a priority.

A further check on the quality of the structure of *Tp.acidophilum* was carried out using the program ENVIRONMENTS (Luthy *et al.*, 1992). This program assesses the three-dimensional environment of the individual residues in a structure and scores them depending on how acceptable are their environments. Fig. 8.9 shows the 3D-1D profile for a single monomer of *Tp.acidophilum* citrate synthase. Only the C-terminus and residue 152 are in unacceptable environments. As mentioned previously, the C-terminus needs more attention in manual rebuilding and refinement and residue 152 is an intersubunit contact residue and therefore may be forced in an unfavourable environment. Nevertheless, the majority of the partially refined structure lies in an acceptable environment as determined by the 3-dimensional profile plot. A profile plot for the dimer (Fig. 8.10) reveals that the bad score (ie. below zero) for residue 152 is probably a function of the monomer being analysed instead of the functional dimer.

The overall stereochemistry of the model at the present stage of refinement is acceptable with RMS deviations of bond lengths and bond angles of 0.022Å and 4.50° respectively.

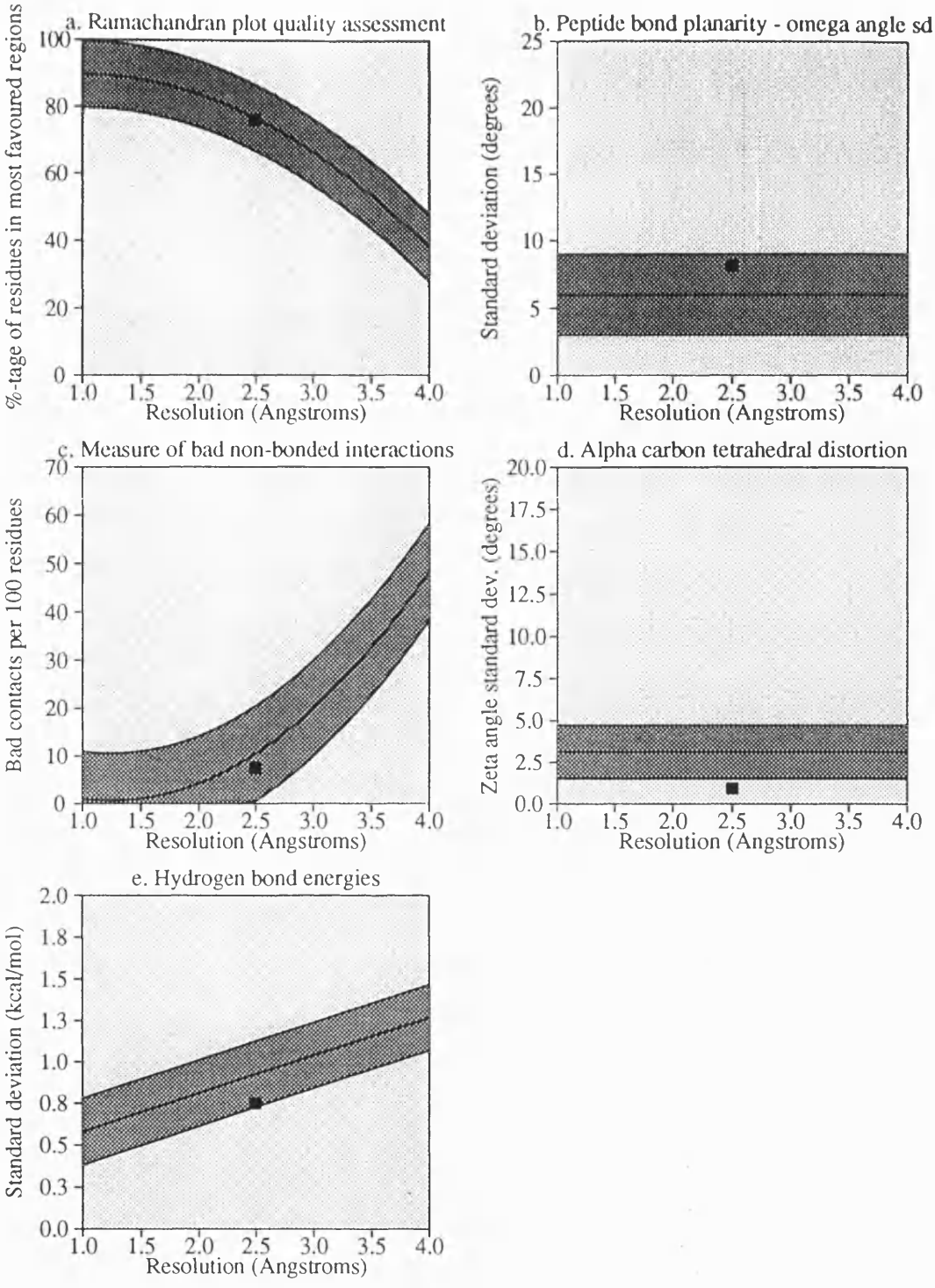


Figure 8.7: Analysis of main chain parameters for a monomer of the structure of *Tp.acidophilum* citrate synthase.

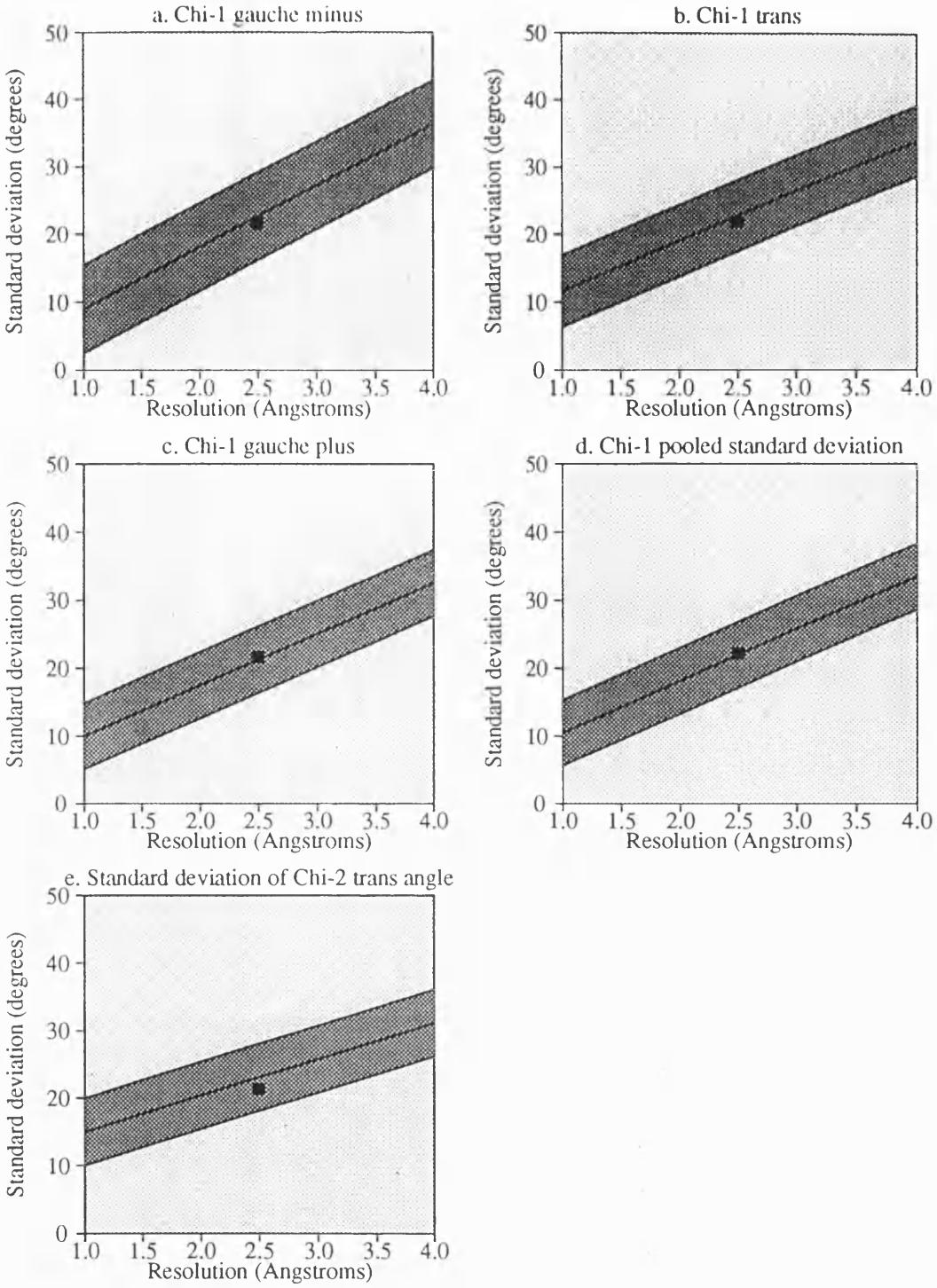


Figure 8.8: Analysis of side chain parameters for a monomer of the structure of *Tp.acidophilum* citrate synthase.

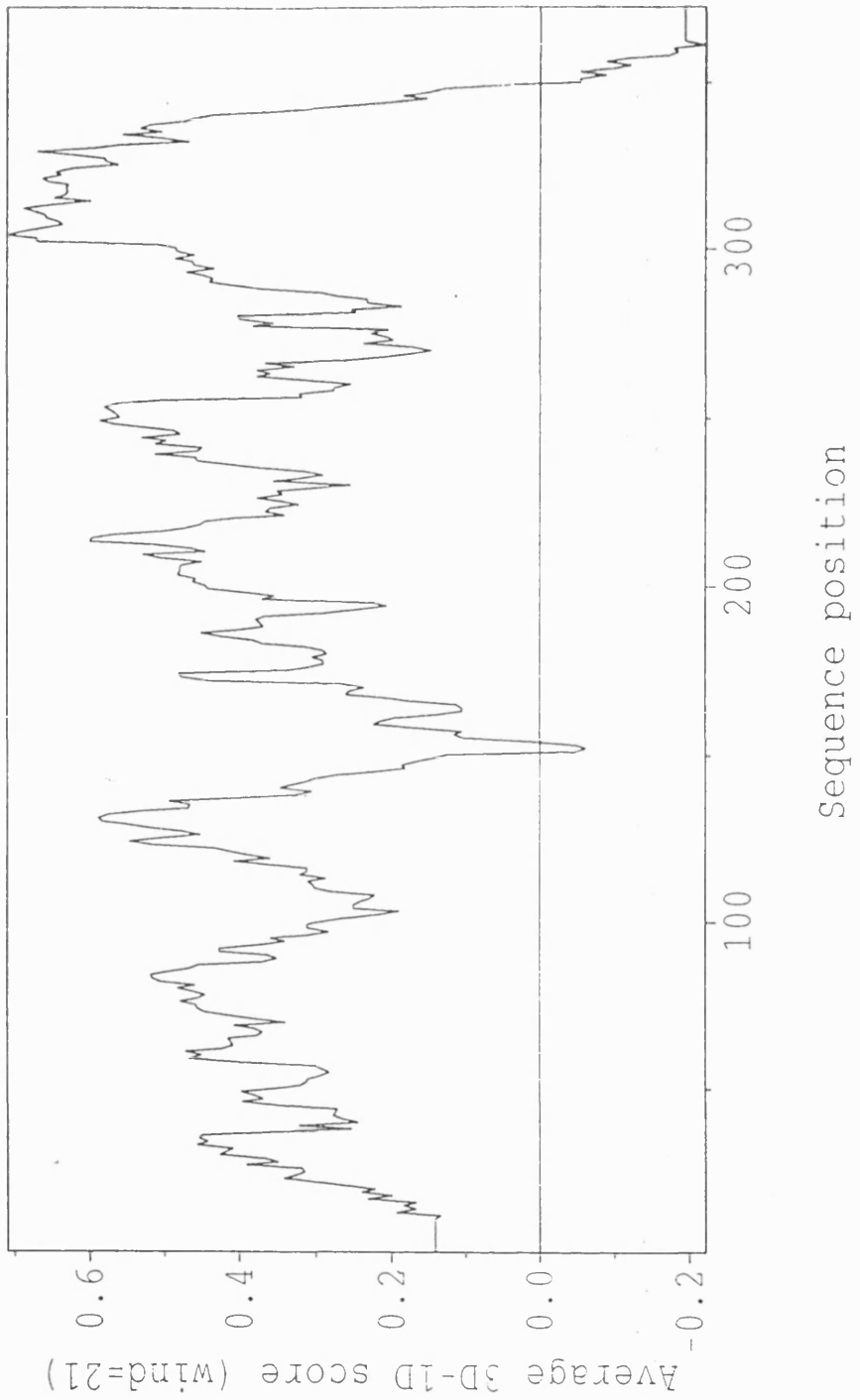


Figure 8.9: 3D-1D profile of a monomer of the partially refined structure of *Tp.acidophilum* citrate synthase.

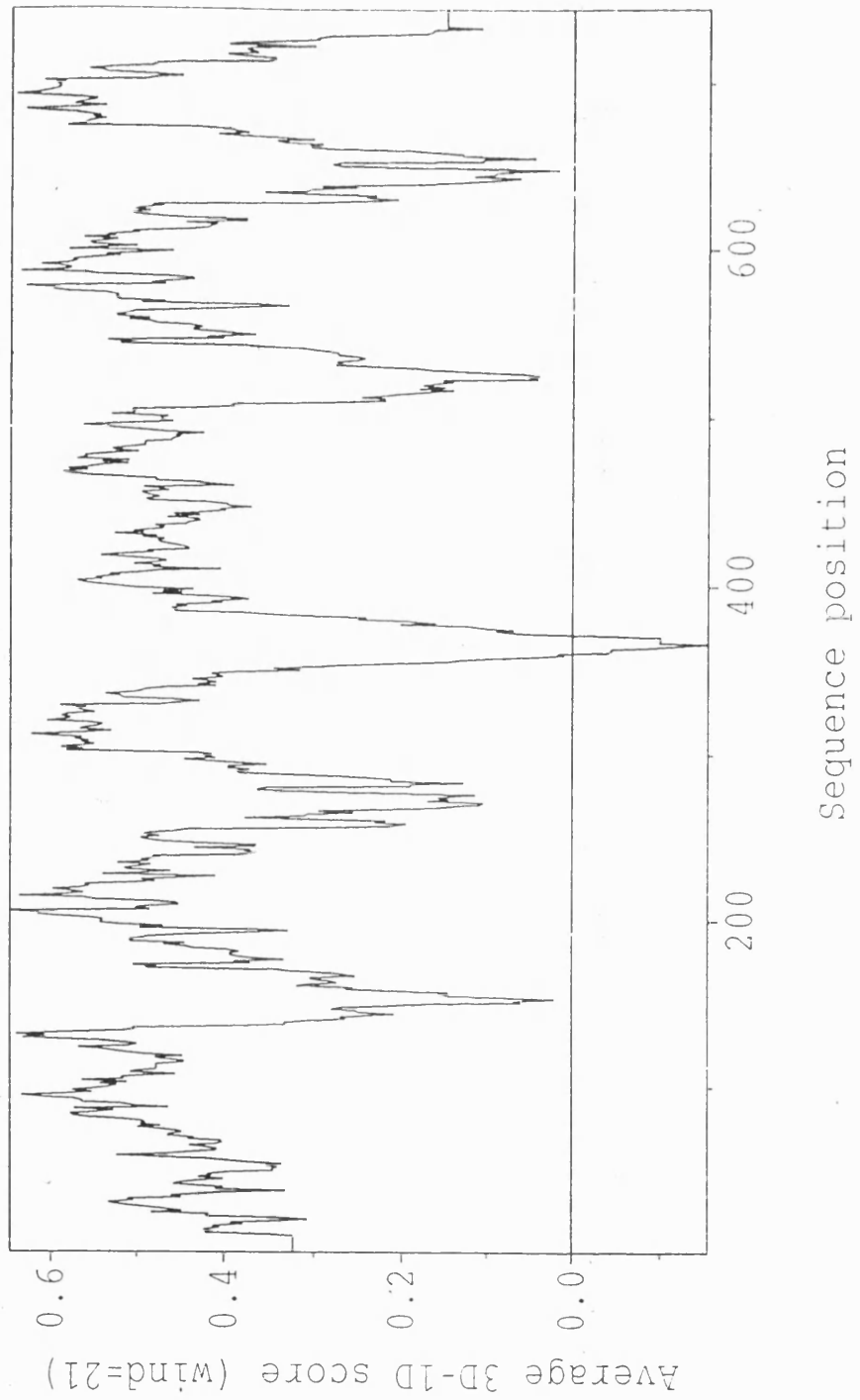


Figure 8.10: 3D-1D profile of a dimer of the partially refined structure of *Tp.acidophilum* citrate synthase.

Chapter 9 : Structural characterisation of *Thermoplasma acidophilum* citrate synthase

9.1 INTRODUCTION

The partially refined crystal structure of *Tp.acidophilum* citrate synthase is only the second reported 3-D structure of an archaeal enzyme and as such is of great interest with respect to highlighting features that may be archaeal in nature and ones that may account for the thermostability of the enzyme. Comparative structural studies may also give an insight into the evolutionary relationships between the Archaea and other living organisms by identifying possible primitive forms of enzymes. The open form of pig heart citrate synthase, which has been refined to 2.7 Å (Remington *et al.* 1982) has been used for all comparative studies.

9.2 OVERALL FOLDED CONFORMATION

Tp.acidophilum citrate synthase is composed of a dimer of identical monomers - Fig. 9.1 and 9.2, with the monomers related by a 2-fold operation. Each monomer consists of 19 α-helices (A-R), equivalent to 60% of the residues with helical secondary structure (as predicted by DSSP (Kabsch and Sander, 1983)) - Table 9.1 gives a breakdown of secondary structural elements and their equivalent elements in pig heart citrate synthase. This helical value correlates well with the helical content predicted by the circular dichroism studies (see Chapter 4). There is a small region of anti-parallel β-sheet (residues 21-23 and 30-32) located towards the N-terminus which forms a tight hairpin loop. The β-sheet conformation region in the Ramachandran plot (Fig. 8.2) has a significant population due to the extended nature of the loops connecting the α-helical regions of the protein, nearly all of which are solvent accessible. The monomer can

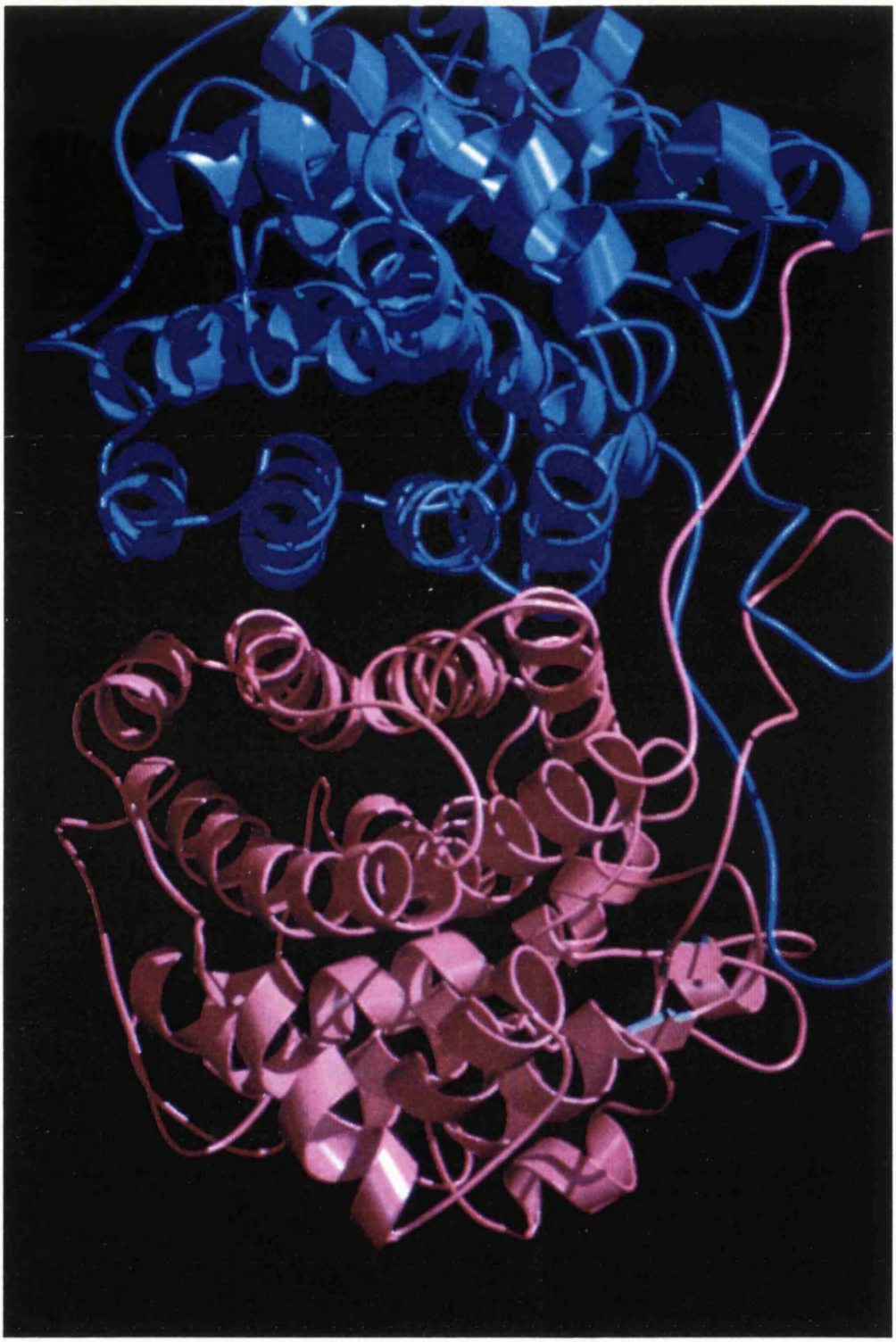


Figure 9.1: Schematic representation of the dimer of *Tp.acidophilum* citrate synthase. Created using MOLSCRIPT (Kraulis, 1991) and RASTER3d (Bacon *et al.*, 1988).

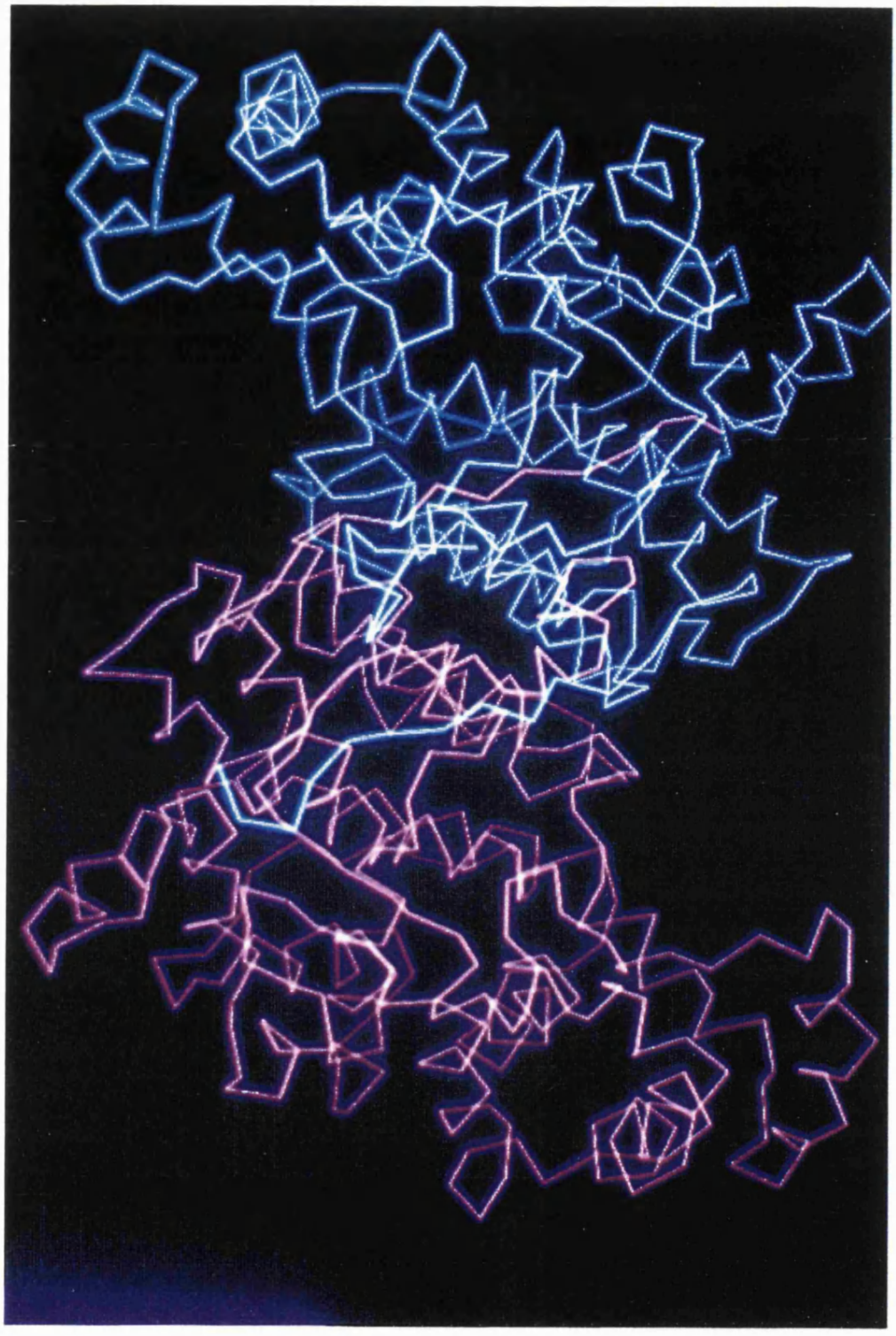


Figure 9.2: α-Carbon backbone of the dimer of *Tp.acidophilum* citrate synthase, viewed down the two-fold axis. One monomer is coloured cyan, the other magenta. Created using O (Jones *et al.*, 1991).

Thermoplasma	N-term	C-term		Pig heart	N-term	C-term	
				A	7 LYS	27 GLN	α -helix
				B	38 VAL	42 TYR	α -helix
	21 THR	23 ILE	ext. strand		57 VAL	59 ASP	ext. strand
	30 LEU	32 TYR	ext. strand		63 GLY	65 ARG	ext. strand
A (C)	37 VAL	42 ALA	α -helix	C	71 ILE	77 MET	α -helix
B (D)	48 GLU	55 LEU	α -helix	D	89 PRO	97 VAL	α -helix
C (E)	63 GLN	73 GLN	α -helix	E	104 GLU	117 ARG	α -helix
D (F)	80 ASP	87 ARG	α -helix	F	122 SER	129 ASP	α -helix
E (G)	95 ALA	106 MET	α -helix	G	137 PRO	146 THR	α -helix
				H	153 ASN	160 GLU	α -helix
F (I)	120 ASP	142 ILE	α -helix	I	167 TYR	193 LEU	α -helix
G (J)	158 GLU	166 GLY	α -helix	J	209 TRP	217 LEU	α -helix
H (K)	172 GLU	184 TYR	α -helix	K	222 ALA	234 ILE	α -helix
I (L)	191 ALA	201 SER	α -helix	L	243 VAL	250 LEU	α -helix
J (M)	206 MET	217 LEU	α -helix	M	258 PRO	269 LEU	α -helix
K (N)	225 ALA	236 ILE	α -helix	N	278 ASN	291 GLU	α -helix
L (O)	241 MET	249 ASN	α -helix	O	300 LYS	310 ASN	α -helix
M (P)	270 PRO	278 ILE	α -helix	P	328 PRO	340 HIS	α -helix
N (Q)	288 GLU	291 LYS	α -helix	Q	345 PRO	363 GLU	α -helix
O ( )	294 GLU	299 LEU	α -helix	R	375 ASP	383 GLN	α -helix
P ( )	302 PHE	311 GLY	α -helix	S	393 TYR	411 ALA	α -helix
Q (R)	316 THR	327 ILE	α -helix	T	427 THR	434 VAL	α -helix
R ( )	335 ILE	357 GLU	α -helix				
S ( )	374 GLU	377 TYR	α -helix				

Table 9.1: Secondary structural breakdown of a monomer of *Tp.acidophilum* citrate synthase and pig heart citrate synthase as predicted by DSSP (Kabsch and Sanders, 1982).

be divided into two domains, a small domain comprising helices KLMNOPQ (residues 225-327 approximately) and a large domain. The subunit-subunit interface consists of an 8 α-helix sandwich, made up of 4 anti-parallel pairs of helices DD', EE', II' and JJ'. The axes of these helices are approximately perpendicular to the 2-fold axis. Wrapped around this core of the dimer are a pair of anti-parallel helices F and R both of which bend smoothly. This bending is present in other α-helices throughout the molecule, and was also observed in pig heart citrate synthase. Helix I in pig heart citrate synthase is bent due to the presence of Pro183 in the middle of the helix, but the corresponding helix in *Tp.acidophilum* citrate synthase (helixF) has no internal proline residue but still forms a bent helix (possibly due to a intra-helical GLY), but to a lesser extent. Therefore the bending of this pair of helices must be necessary for the correct conformation and function of citrate synthase.

The α-helices in each monomer are arranged in layers with each layer comprising roughly anti-parallel pairs of helices. This layering has the effect of causing many of the helices to be completely buried from solvent.

9.2.1 Active site conformation

The active site of pig heart citrate synthase is situated in the cleft between the two domains having been identified by substrate/analogue bound crystal structures (Remington, 1992). On binding of OAA, the open form of the enzyme undergoes a conformational change equivalent to an 18° rotation of the small domain with respect to the large domain thereby generating the closed form of the enzyme.

From sequence alignment studies 8 out of 11 essential catalytic active site residues in pig heart citrate synthase are conserved in *Tp.acidophilum* citrate synthase, with two of the remaining residues showing a conserva-

tive change (ARG–>LYS). The active site of *Tp.acidophilum* citrate synthase shows a high degree of similarity to pig heart citrate synthase - see Fig. 9.3. Three residues have been implicated as essential for catalysis – ASP375, HIS274 and HIS 320 (pig heart residues). *Tp.acidophilum* citrate synthase contains three structurally homologous residues – ASP317, HIS222 and HIS262, in a very similar conformation. HIS274 is conformationally strained (Remington *et al.*, 1982) in pig heart citrate synthase but not in *Tp.acidophilum*, possibly due to a slight shift of the position of this residue in the latter towards the active site pocket. Two GLY residues are the next residues after this catalytic HIS and the flexibility of GLY residues may allow the HIS to occupy a more conformationally favourable orientation. Three ARG residues (329, 401 and 421 from the other monomer) and two HIS residues (238 and 320) have been implicated as necessary for citrate or oxaloacetate binding. All these residues have their structural equivalents in *Tp.acidophilum* citrate synthase – ARG 271, 344 and 364 and HIS 187 and 262. Pig heart citrate synthase has three ARG residues (65 and 324 and 164 from the other monomer) capable of acting as ligands in the binding of CoenzymeA, as well as H-bonds from main chain residues 315-318. In *Tp.acidophilum* citrate synthase ARG65 has a structural equivalent in ARG31, but ARG324 has been replaced by LYS266 (as predicted from sequence alignments). However, ARG164 has no structural equivalent residue (as predicted by SHP) it residing in a surface loop which has been deleted in the thermophilic structure. LYS114, although predicted as equivalent to another residue, probably is able to take the role of acting as the third ligand in binding of the three phosphate groups, although in the absence of a closed form structure this is hard to verify.

The overall folded conformations of *Tp.acidophilum* and pig heart

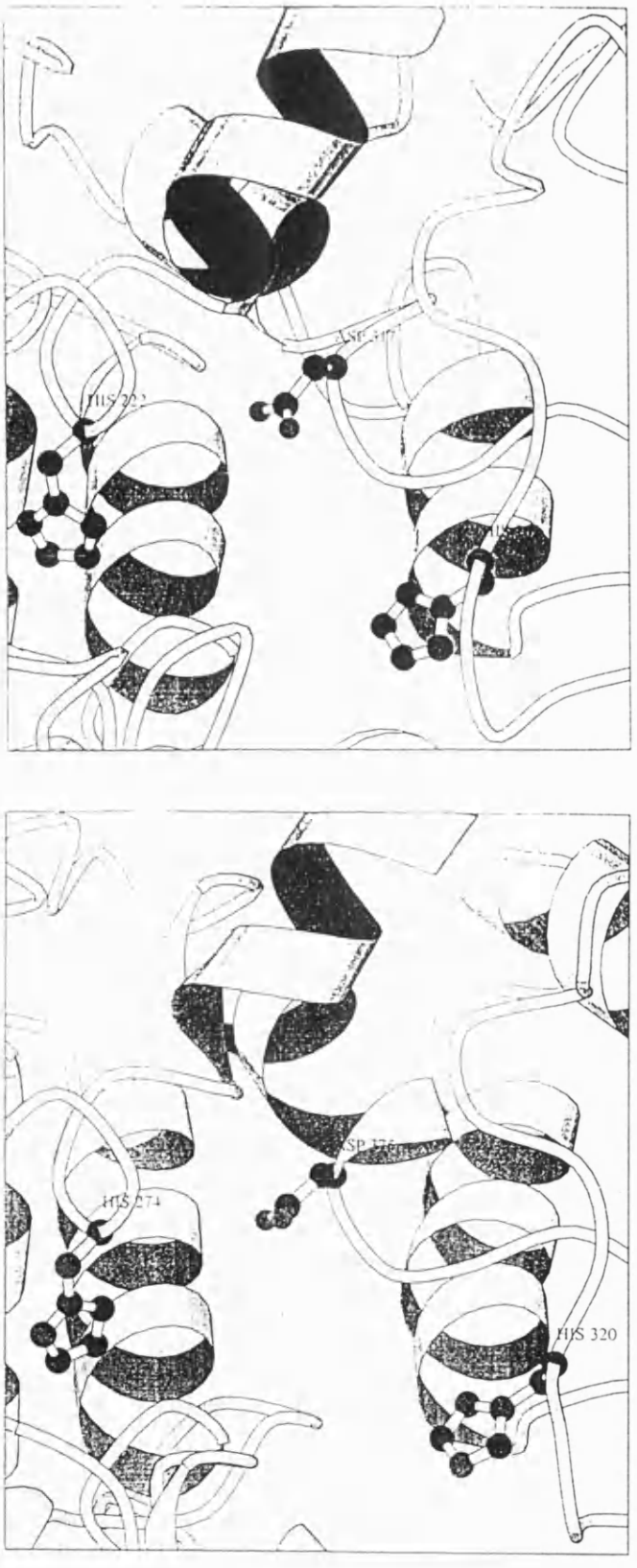


Figure 9.3: Active site conformation of *Tp.acidophilum* (top) and pig heart (bottom) citrate synthases. Only three catalytically vital residues are highlighted for simplicity. Created using MOLSCRIPT (Kraulis, 1991).

citrate synthases are very similar - see Fig. 9.4. The RMS distance between the two enzymes is 2.15, with 363 structural equivalences (SHP (Stewart, 1979)) – equivalent to 95% of the former and 83% of the latter. The core of the protein shows the highest degree of similarity, with more divergence occurring towards surface regions. In pig heart citrate synthase the small domain consists of 5 helices NOPQR (residues 278-383 approximately), whereas *Tp.acidophilum* citrate synthase consists of 7 shorter helices. Although the presence of these extra helices could be due to the stringent nature that DSSP predicts residues that exist in helices and that this is not a fully refined structure. The overall number of amino acids present within the small domain is the same in both enzymes. The position of the helices in the small domain show the largest change within the secondary structural elements, in particular HelixL in *Tp.acidophilum* (equivalent to HelixO in pig heart). This helix is situated at the extreme of the molecule and has been proposed to show relatively high flexibility in pig heart citrate synthase and to be deformed during the reaction pathway of the enzyme (El-Kettani *et al.* 1993).

The structural similarity between a citrate synthase from pig heart and the thermophilic Archaeon *Tp.acidophilum* is striking considering the evolutionary distance between the two organisms. The Archaea are thought to be evolutionary primitive organisms; therefore, *Tp.acidophilum* citrate synthase can be seen as an ancestral form of citrate synthase, with differences in the pig heart structure being possible adaptations away from thermophilicity towards mesophilicity.

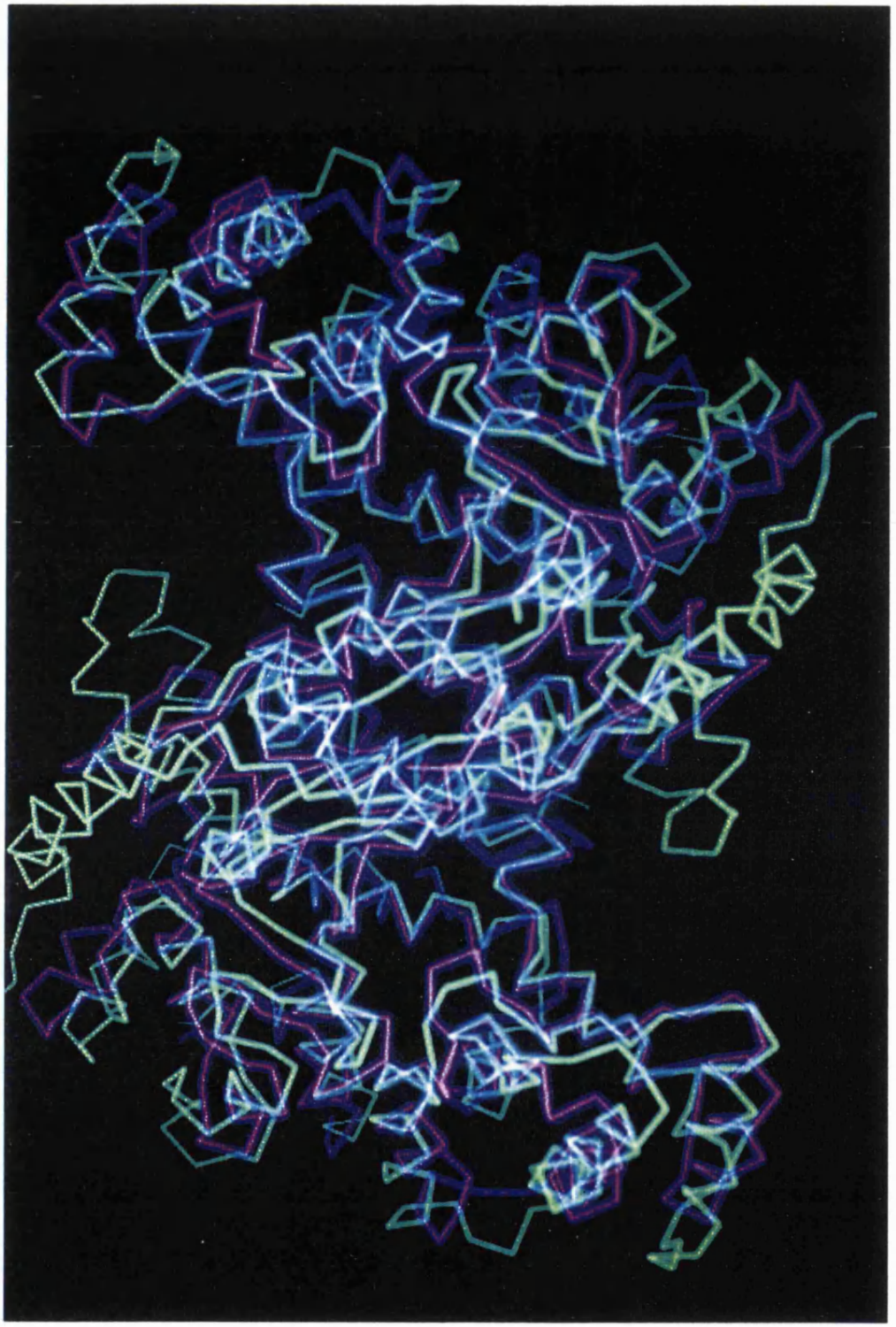


Figure 9.4: Least squares superposition of pig heart citrate synthase (green) onto *Tp.acidophilum* citrate synthase (magenta). Viewed down the 2-fold axis. Created using O (Jones *et al.*, 1991).

9.3 POSSIBLE FEATURES CONFERRING THERMOSTABILITY

9.3.1 N-terminal deletion

Tp.acidophilum citrate synthase consists of 384 residues compared to the 437 residues of pig heart citrate synthase; this is a 12% reduction in size, 68% of which can be accounted for by a 36 residue deletion at the N-terminus. In pig heart citrate synthase these residues form a long α-helix which loops around the surface of the molecule (Fig. 9.5). Pig heart citrate synthase and *Tp.acidophilum* citrate synthase have similar kinetic properties and perform the same function in the cell and so the purpose of this extension is still unknown. The N-terminal deletion is seen in all the archaeal citrate synthases and a thermotolerant Bacillus citrate synthase but not in any others thus far sequenced (Section 1.5.3). Thus this N-terminal deletion may be necessary for adaption to extreme environments and if present may act as a weak point in the thermal unfolding pathway of the mesophilic citrate synthases. This hypothesis is currently being tested in Bath by the creation of a N-terminal deletion mutant of pig heart citrate synthase which will be tested for its kinetic properties and thermostability compared to the wild type enzyme.

9.3.2 Loop deletions

A striking difference between the enzymes of the thermophile and mesophile is the shortening of a number of loops in *Tp.acidophilum* citrate synthase - see Table 9.2 and Fig. 9.6. The reduction in loop sizes was also observed in the crystal structure of a thermophilic phosphoglycerate kinase when compared to its mesophilic counterpart (Davies *et al.* (1992)). Another feature of *Tp.acidophilum* citrate synthase compared to pig heart citrate synthase is the deletion of HelixH (residues 153-160).

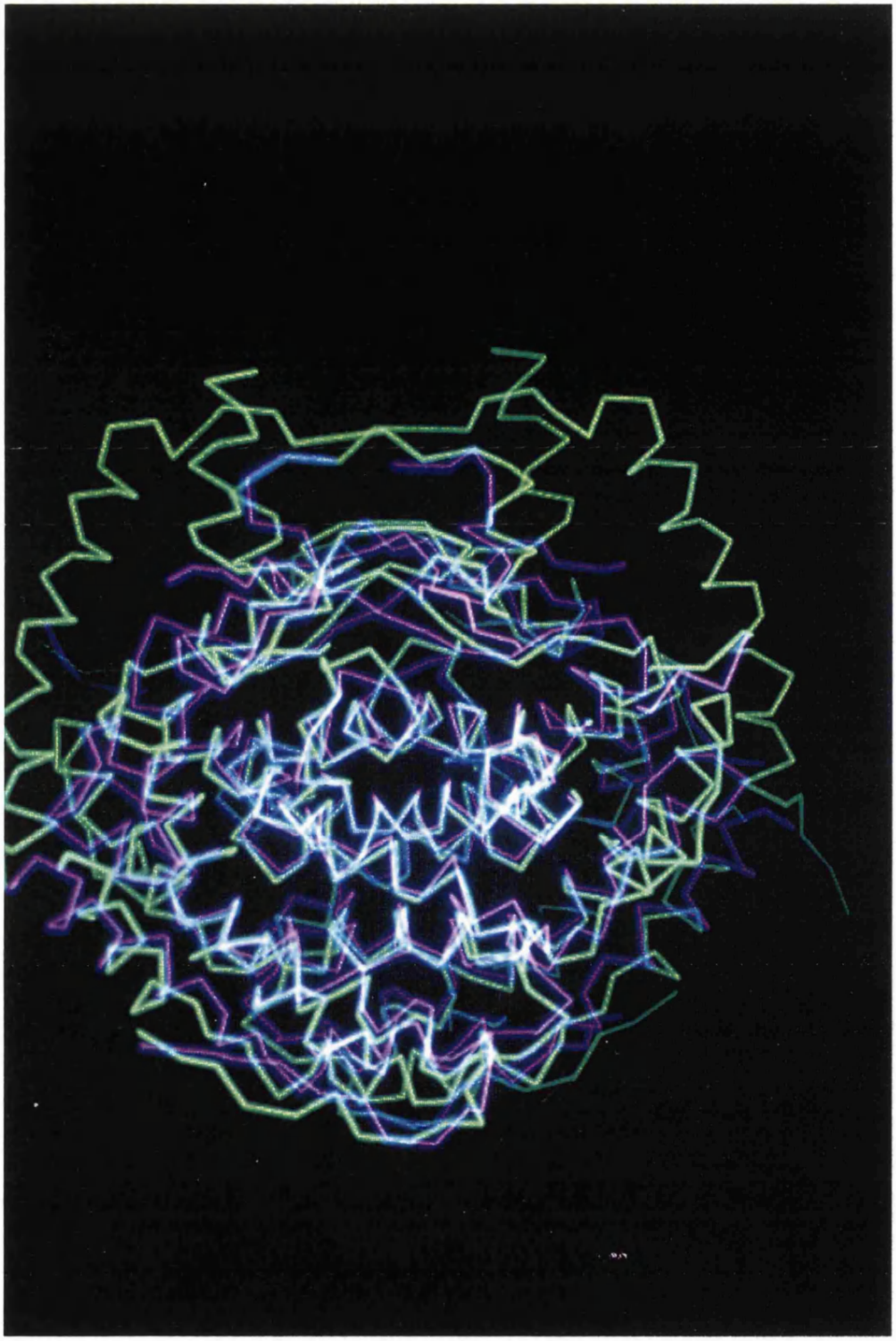


Figure 9.5: Least squares superposition of pig heart citrate synthase (green) onto *Tp.acidophilum* citrate synthase (magenta). Created using O (Jones *et al.*, 1991).

Thermoplasma residues	Pig residues	Pig α-helix deletion
—	1–36	helixA
43–47	78–88	—
107–119	147–166	helixH
237–240	292–299	—
312–315	364–374	—

Table 9.2: Breakdown of loops present in *Tp.acidophilum* citrate synthase, which have been shortened with respect to pig heart citrate synthase.

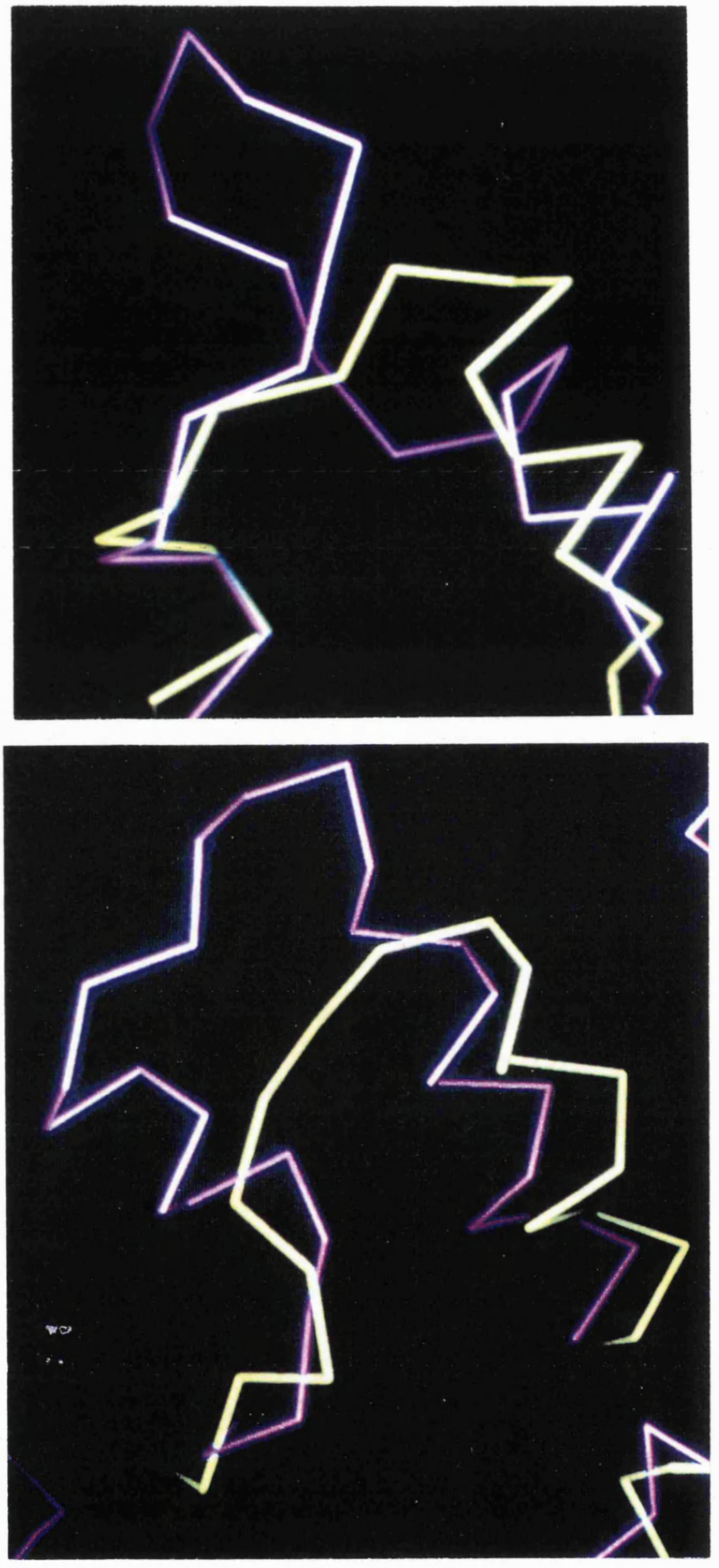


Figure 9.6: Examples of two loop deletions observed in *Tp.acidophilum* citrate synthase (yellow) with respect to pig heart citrate synthase (magenta). Created using O (Jones *et al.*, 1991).

This deletion coincides with the reduction in the size of one of the aforementioned loops in *Tp.acidophilum* citrate synthase. Crystallographic B-factors (Temperature factors) indicate regions of relative flexibility in the crystal structure - high B-factors indicate high relative flexibility. As the temperature increases so will the flexibility and the most flexible regions at normal working temperatures could be the regions that may start to unfold first as the temperature increases. In pig heart citrate synthase there are three distinct regions of flexibility, as judged by a plot of the B-factors of Cα and Cβ atoms; the N-terminal extension and two surface loop regions (78-88 and 292-299) , all of which have been deleted in the *Tp.acidophilum* citrate synthase structure. Therefore this reduction in the loop sizes in the thermophilic enzyme may be necessary to avoid regions of high flexibility and therefore possible weak points for thermal denaturation. A B-factor plot (see Chapter 8) of the Cα atoms of *Tp.acidophilum* citrate synthase shows no significant regions of high flexibility relative to any other region in the molecule. This may be indicative of this structure being solved from crystals grown at $25°C$, a temperature at which the thermophilic enzyme is under sub-optimal conditions and may be less flexible (see Chapter 5).

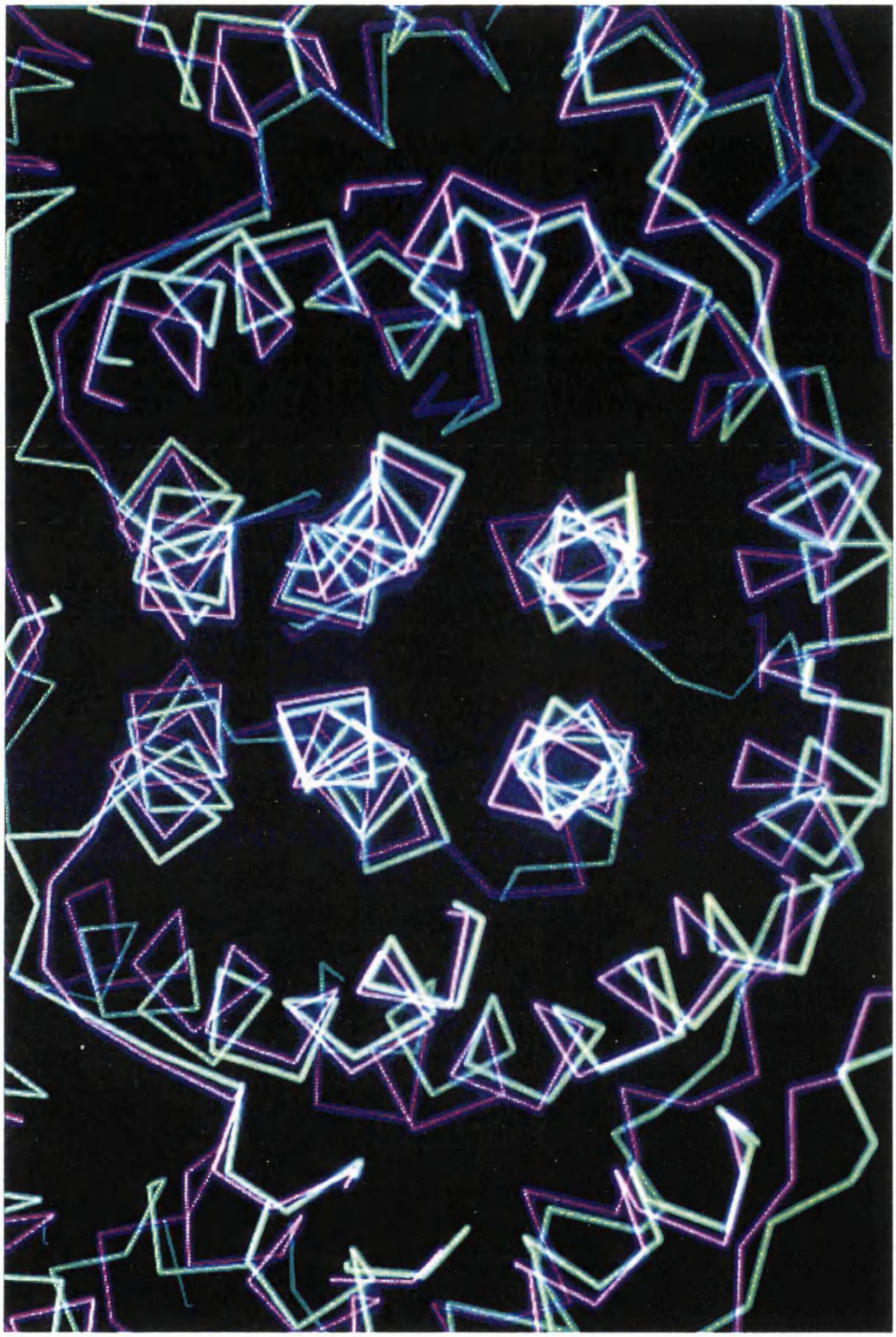
9.3.3 Subunit-subunit interactions

A feature that has been widely studied is whether there are specific amino acid changes that occur between mesophiles and thermophiles (Menendez-Arias and Argos, 1989). One such feature is the increase in alanine content in proteins from thermophiles with respect to their mesophilic counterpart (eg. MDH from *Thermus flavus*). In *Tp.acidophilum* citrate synthase 12% of the amino acids are alanine, 8% are isoleucine and a 6% are leucine compared to 8%, 4% and 12% respectively in pig heart citrate synthase. Although this kind of feature can be identified through sequence compar-

isons, the elucidation of the crystal structure allows the positioning of specific amino acid replacements.

Tp.acidophilum citrate synthase is only active as a dimer of identical monomers and therefore subunit/subunit interactions must be sufficiently strong to keep the enzyme in its native active conformation at its working temperature. Pig heart citrate synthase has an identical subunit/subunit interface conformation: an 8 α-helix sandwich, made up of 4 anti-parallel pairs of helices with 4 helices from each monomers (Fig. 9.7). Upon raising the temperature pig heart citrate synthase has been proposed to dissociate into monomers which then undergo denaturation (McEvily and Harrison 1986). Therefore, differences in amino acids involved the subunit interface must account for the difference in dimer dissociation temperature. In pig heart citrate synthase 21% of the surface of the monomer becomes inaccessible to solvent upon dimerization, whereas only 17% of the *Tp.acidophilum* monomer becomes buried in this process (DSSP). The burial of hydrophobic areas has been shown to be an important factor in protein stability (Fersht *et al.*, 1993). Imada *et al.* (1991) have shown that a thermophilic isopropylmalate dehydrogenase has an increased hydrophobic area in the dimer interface compared to its mesophilic counterpart.

To analyse the differences in subunit/subunit interactions the residues involved in *Tp.acidophilum* citrate synthase were identified, based on the residues involved in pig heart citrate synthase - see Table 9.3. The most significant difference in the contribution of individual amino acids is the increase from 7 alanines involved in pig heart citrate synthase to 15 alanines involved in *Tp.acidophilum* citrate synthase. There is also a marked reduction in the number of leucine and glycine residues involved in the subunit/subunit contacts, and a complete absence of histidine,



(Figure 9.7): Subunit-subunit interface of both pig heart (green) and *Tp.acidophilum* (magenta) citrate synthase. Created using O (Jones *et al.*, 1991).

Pig residue	Equivalent Thermoplasma residue	Pig residue	Equivalent Thermoplasma residue
GLY 29	—	VAL 38	THR 3
GLY 34	—	VAL 52	TRP 17
GLY 44	LYS 8	VAL 57	THR 22
GLY 50	ASN 14	VAL 243	ALA 191
GLY 161	THR 115	VAL 251	ALA 199
GLY 241	VAL 189	VAL 314	ARG 256
GLY 268	ALA 216	VAL 322	VAL 264
GLY 271	GLY 219		
GLY 312	GLY 253	ASN 130	GLN 88
		ASN 134	GLU 92
LEU 33	—	ASN 149	ALA 107
LEU 51	ILE 15	ASN 267	ALA 215
LEU 135	SER 93		
LEU 163	—	ILE 21	—
LEU 250	VAL 198	ILE 36	—
LEU 255	LEU 203	ILE 431	—
LEU 260	SER 208		
LEU 273	LEU 221	PRO 132	PRO 30
LEU 276	GLY 224	PRO 272	PRO 220
LEU 419	LEU 362	PRO 418	—
LEU 430	—	PRO 422	—
LEU 433	—		
		ALA 32	—
HIS 123	PHE 81	ALA 143	ALA 101
HIS 136	ASP 136	ALA 147	ALA 105
HIS 246	THR 194	ALA 254	THR 202
		ALA 264	ALA 212
LYS 49	ASN 14	ALA 270	THR 194
LYS 423	ARG 366		

Table 9.3: Residues involved in subunit-subunit contacts in both pig heart and *Tp.acidophilum* citrate synthases.

Pig residue	Equivalent Thermoplasma residue	Pig residue	Equivalent Thermoplasma residue
		PHE 24	—
MET 45	LEU 10	PHE 131	LEU 89
MET 48	VAL 13	PHE 415	GLU 358
MET 127	ALA 85		
MET 425	VAL 368	TYR 42	SER 7
		TYR 158	—
GLN 35	—		
		GLU 54	ARG 19
ARG 20	—	GLU 239	GLU 188
ARG 25	—	GLU 280	ALA 228
ARG 46	GLU 11	GLU 420	ILE 363
ARG 164	—		
ARG 313	LYS 255	SER 41	—
ARG 421	ARG 364	SER 139	ALA 97
		SER 150	ALA 107
THR 37	—	SER 256	SER 204
THR 126	ASN 84	SER 424	ALA 367
THR 146	ALA 104	SER 426	—
THR 247	ALA 195		
THR 427	—	ASP 39	GLU 4
		ASP 428	—

Table 9.3: Contd.

Amino Acid	PIG	THERMOPLASMA
GLY	9	3
ALA	6	15
ASN	4	2
HIS	3	0
LYS	2	2
LEU	12	5
VAL	7	6
ILE	3	2
PRO	4	2
PHE	3	1
MET	4	0
TYR	2	0
ASP	2	1
ARG	6	4
SER	6	4
GLN	1	1
GLU	4	5
THR	5	5
TRP	0	1

Table 9.4 Number of specific amino acids involved in subunit interface in pig heart and *Tp.acidophilum* citrate synthase.

methionine and tyrosine residues. The amino acid changes highlighted here could be indicative of the global changes in amino acid composition. To analyse the distribution of these changes, helical wheels (GCG, Devereux *et al.*, 1984) were performed on the 4 α-helices from each monomer in the interface. Helical wheels are useful tools in investigating the hydrophobic/ hydrophilic nature of α-helices. Helices FLM in pig heart citrate synthase are hydrophobic on the side facing the other monomer (see Fig. 9.8) but helix G is hydrophilic, whereas in *Tp.acidophilum* citrate synthase the equivalent helices DEIJ all have hydrophobic inter-subunit faces. The major change in amino acid composition between helix E (in *Tp.acidophilum* citrate synthase) and helix G (in pig heart citrate synthase) is the increase in alanine content (Fig. 9.8): 8 in the former and 3 in the latter, a larger change than would be expected by the global amino acid changes. Horovitz *et al.* (1992) have shown that alanine residues in α-helices are the most stabilising amino acid due to decreased flexibility and increased hydrophobicity. O'Neil and DeGrado (1990) investigating synthetic small peptide stability also showed the stabilising effects of alanine in α-helices. The methyl side-chain of alanine is thought to reduce the conformational freedom of the main chain in the unfolded state, thus leading to a less unfavorable change in entropy upon helix formation. Residues that are able to form H-bonds with the main chain in the unfolded state tend to be helix destabilizing. Two of the observed changes in helixE of *Tp.acidophilum* citrate synthase correspond to a SER->ALA substitution which has been identified as the second most common substitution in mesophile -> thermophile. O'Neil and DeGrado have shown that this substitution at an internal position causes at 0.4Kcalmol^{-1} increase in the free energy of the folded state. Menendez-Arias and Argos (1989) reported there are two regions in which

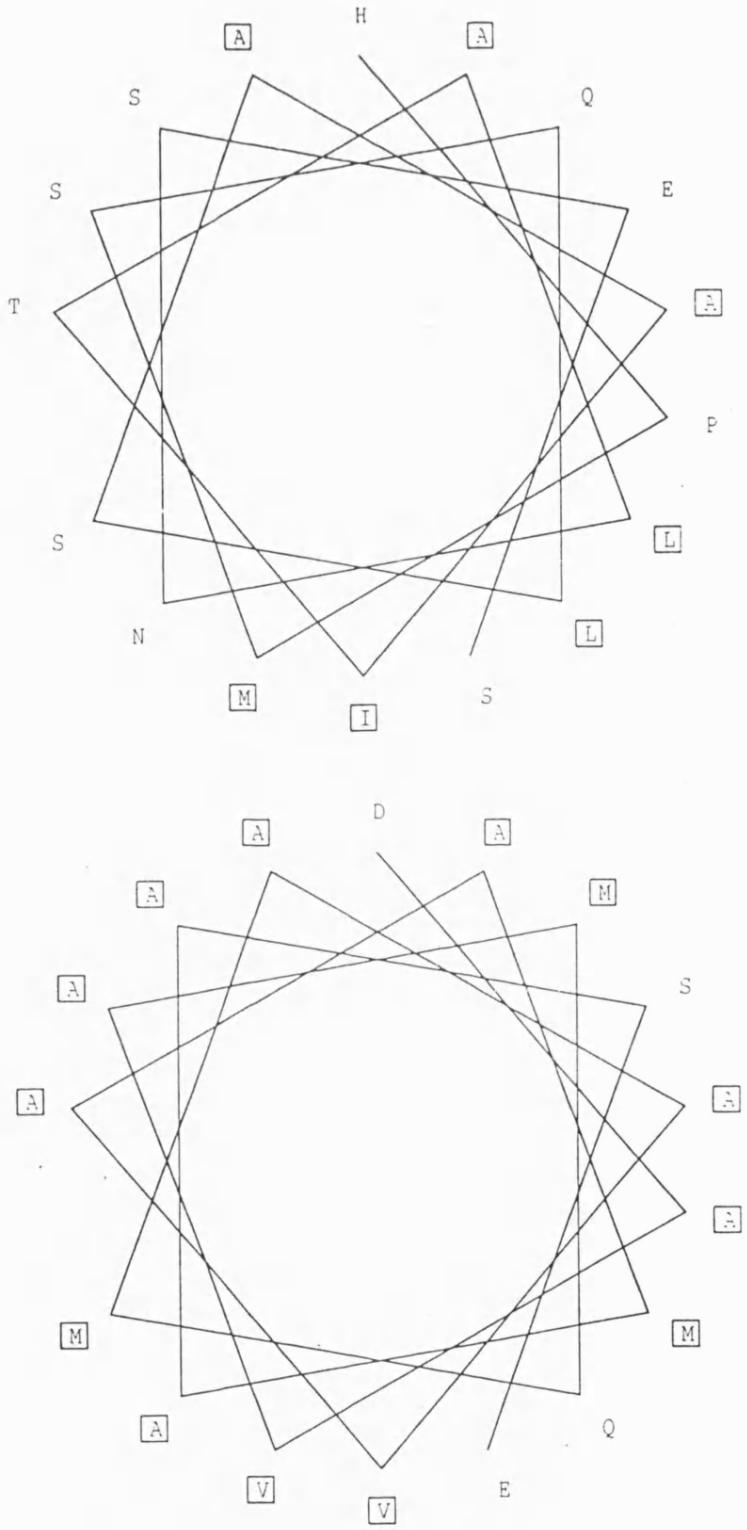


Figure 9.8: Helical wheels (GCG) of Helix G of pig heart citrate synthase (top) and Helix E of *Tp.acidophilum* citrate synthase (bottom). The left hand face of each helix resides in the subunit interface. Boxed residues are hydrophobic in nature.

decreased flexibility and increased hydrophobicity correlate well with thermostability, namely α-helices and to a lesser extent domain interfaces. Kotik and Zuber (1993) recently reported an impressive $20°C$ stabilization of lactate dehydrogenase from *Bacillus megaterium* due to 2 specific mutations in the subunit interface. A serine and a threonine in α-B helix were mutated to alanines. These mutations also resulted in a decreased overall flexibility of the enzyme. They propose that the compo- sition of α–helices of structural importance (ie. subunit interfaces) may be critically important in stabilization of a protein. Helix E in *Tp.acidophilum* citrate synthase is both an α-helix and is part of the subunit/subunit interface indicating that this may be a specific feature of *Tp.acidophilum* citrate synthase that may be responsible for its thermostability.

9.3.4 Helix-capping residues

α-Helices have an inherent macro-dipole along their length: a positive charge at the N-terminus and a negative charge at the C-terminus, which may be destabilizing if the charge is not counterbalanced. Helix capping at the N-terminus by negatively charged residues (ie. ASP and GLU) or neutral polar residues (ie. SER or THR) has been proposed as a counteractive measure to increase the stability of α-helices (Serrano *et al.*, 1992). Similarly a positively charged residue at the C-terminus is stabilizing. Table 9.1 shows the secondary structural characteristics of the two citrate synthases and also highlights the residues at the N- and C- termini. The thermophilic protein has ten negatively charged residues at the N-termini of its helices compared to only 5 in the mesophilic enzyme. Pig heart citrate synthase has 4 positively charged residue in the Carboxy-cap position, with only three in *Tp.acidophilum* citrate synthase but the thermophile does have 2 helices carboxy-capped by GLY residues which are thought to be the most

stabilizing residue in the C-cap position.

Harper and Rose (1993) have extended this concept towards a helix-capping box motif. The first three residues in an α-helix are not involved in the standard hydrogen-bonded network seen in the centre of an α-helix , leaving atoms capable of H-bonding without a partner. An extensive survey of known protein structures studying the N-terminal regions of α-helices, revealed a hydrogen-bonding pattern where the N-cap residue forms a H-bond with the backbone >NH of N3 and, reciprocally the side-chain of N3 forms a H-bond with the backbone >NH of Ncap. The residue preceding the start of the helix (N-cap) has phi/psi angles of 94°+/-15° and 167°+/-5° respectively, outside those for α-helical residues. The capping box can be generalised into an S-X-X-E motif, where S is the N-cap residue and E is the N3 residue. There is strong residue preference for the capping box motif. Both Ncap and N3 must be hydrophilic and stereochemically capable of forming the hydrogen-bonded pattern (eg. SER,THR,GLU and ASP). The most abundant normalized pairwise frequency of occurrence is SER-GLU.

In their extensive survey they identified pig heart citrate synthase as having two observed helix capping box motifs in helices A and C and one potential one in helix J (defined as lacking one of the complementary H-bonds). On the other hand, *Tp.acidophilum* citrate synthase has 4 potential helix capping boxes in helices A,B,C and J (Table 9.5). The boxes being potential may be due to the structure being only partially refined. Pig helix C and *Tp.acidophilum* helix A are equivalent structural helices. Of all the proteins surveyed the greatest number of observed and potential helix capping boxes is three: in pig heart citrate synthase and D-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus steareothermophilus*.

Thermoplasma residues	Helix	Pig residues	Helix
S-V-E-D	A	T-V-A-M	B
Q-D-E-E	B	S-I-P-E	C
T-E-Q-E	C	D-W-S-H	J
D-M-Y-S	J	—	–

Table 9.5: Residues proposed to be involved in helix capping box motifs in *Tp.acidophilum* citrate synthase.

Pig heart citrate synthase contains 20 α-helices (60% of residues) and therefore the three helix-capping boxes may be a function of the number of helices present. *Bacillus steareothermophilus* is a thermophilic bacterium that may be using the helix capping box for stabilising the α-helices present in D-glyceraldehyde-3-phosphate dehydrogenase, since this enzyme contains only 8 α-helices (25% of residues). *Tp.acidophilum* citrate synthase has four helix capping boxes three of which are situated in the first three helices at the N-terminus; this compares to 2 N-terminally-situated helix capping boxes in pig heart citrate synthase - see Table 9.4. The stabilization of the first three α-helices in *Tp.acidophilum* citrate synthase by the presence of a helix capping box may be necessary for activity at the enzyme's elevated working temperature.

9.3.5 Active site conformation/environment

Zhi *et al.* (1991) have created active site mutants of pig heart citrate synthase which reduced the enzymic specific activity by 10^3–10^4 fold but increased the thermostability by up to 10°C, eg ASP375 -> GLN. The amino acid substitutions that increased the thermal stability were ones that replaced charged active site residues with more hydrophobic residues, leading to an increased net positive charge at the active site. The results suggest that specific electrostatic interactions in the active site of pig heart citrate synthase can increase its conformational stability. Mutation studies by Meiering *et al.* (1992) have shown that the active site of Barnase has not evolved to confer the highest stability possible. The active site residues in pig heart and *Tp.acidophilum* citrate synthase are identical, suggesting the same mechanism of action for both enzymes ,although there are subtle differences in the conformation and surrounding residues. HIS274 in pig heart citrate synthase is in a strained, conformationally unfavorable orientation.

Zhi *et al.* (1991) have also proposed that the above mutations may also relax the conformational strain within the active site. In *Tp.acidophilum* citrate synthase HIS274 is not conformationally strained as judged from the Ramachandran plot. To study the electrostatic and hydrophobic environment of the active site of *Tp.acidophilum* citrate synthase with respect to pig heart citrate synthase the residues within a 10 Å sphere of the core of the active site of each the enzymes were determined. The overall charge of this active site sphere was -3 in pig heart but +4 in *Tp.acidophilum* citrate synthase. Therefore the increased positive charge of the environment around the active site may contribute to the enzyme's stability.

9.4 POSSIBLE ARCHAEAL PROTEIN FEATURES

The structure of *Tp.acidophilum* citrate synthase described above is only the second crystal structure of an enzyme from an Archaeon and can be used as a basis for theories about evolutionary ancestral forms of enzymes ie. structural features that may be characteristically archaeal. The most striking feature of the archaeal enzyme is how conformationally similar it is to its counterpart in a mammal. The r.m.s difference over 363 equivalence atoms of 2.15 (SHP program) reveals that the conformation present in the evolutionary old organism must have had the required efficiency for its role and that there were strong evolutionary pressures on maintaining the same overall conformation in evolutionary ephemeral organisms.

The most obvious characteristic of *Tp.acidophilum* citrate synthase is the number of amino acids present compared to mesophilic counterparts. The same overall tertiary conformation is present in both evolutionary distant organisms, but the size reduction is accounted for by deletion of loops and the N-terminus. The evolutionary transition from thermophile to mesophile may have required the reduction in relative stability of enzymes, to

create an enzyme with the same flexibility at the lower temperature, since the enzyme must have sufficient flexibility to function. Therefore the least disruptive way of reducing this stability may have been insertion in loop regions, so that the overall (working) conformation of the enzyme is changed as little as possible. It has been seen in proteins from phylogenetically diverse organisms that the same overall fold has been maintained in both organisms, suggesting that the conformation of proteins was sufficient to perform efficiently their role in the earliest of organisms and that this conformation was maintained throughout evolution.

9.5 CONCLUSION

The above results have highlighted possible areas/reasons for the different stability of the two structurally homologous enzymes, allowing the possibility of rationally designing mutations to alter the thermal stability of proteins. It must be mentioned that the characterisation is not fully completed due to the structure of *Tp. acidophilum* citrate synthase not having been fully refined, but further investigation into specific interactions, eg. hydrophobic and electrostatic, may reveal further structural differences between the mesophilic and thermophilic citrate synthases.

Chapter 10 : Conclusion and Further Work

Structure/function studies on thermostable archaeal proteins allow insights into two distinct areas of scientific interest; namely, the causative factors of thermal stability and identification of putative ancestral forms of proteins. Citrate synthase is a ubiquitous metabolic enzyme ,about which a wealth of primary and tertiary information is known and was therefore was chosen as a model system from which to gain the above insights.

Archaeal organisms are notoriously poor growers in the laboratory, producing low cell masses and therefore low amounts of the protein of interest. Therefore, the previous (cloning, sequencing and) overexpression of the gene for *Tp.acidophilum* citrate synthase in the 'fast-growing' *E.coli* (Sutherland *et al.*, 1990, 1991) has allowed the production of 5-7mg of pure protein from 1l of cells, allowing wide spread crystallization trials.

The elucidation of the crystal structure of *Tp.acidophilum* citrate synthase followed the lines of the old saying; 'If at first you don't succeed try, try again'. Once suitable conditions for crystal growth had been found, three distinct crystal types were identified but quality high resolution data (to 2.4Å) was collected for a single crystal type and from a single crystal. The three data sets enabled the molecular replacement calculations to be carried out on each one. In order to employ molecular replacement techniques, a high level of confidence must be held that the unknown structurally resembles the search model. *Tp.acidophilum* citrate synthase exhibits only a 20% sequence identity to functionally identical pig heart citrate synthase, but sequence alignments show conservation of the proposed active site residues. Circular dichroism studies were also important in that similar levels of secondary structural elements were predicted to exist in both enzymes.

The eventual success of molecular replacement, in this case, has provided some guidelines which may be useful in the solution of other such difficult cases, where a only a poor search model is available. Of paramount importance is the collection of a high quality complete data set. If a solution is not forthcoming with one molecular replacement package, other ones should be tried in conjunction with alternative deletion variations of the original search model. As much information must be gathered concerning the characteristics of the search model. For example, the monomer of pig heart citrate synthase comprises a small and a large domain and therefore in PC-refinement in X-PLOR they can be used as individual rigid bodies to try and enhance possible rotation function peaks. In this way many subtly different search models were able to be investigated.

Structural comparisons between the mesophilic and thermophilic citrate synthases allow identification of possible features that confer the latter's enhanced stability, for example, loop deletions, subunit-subunit interactions and helix-capping residues. Since the structure is only partially refined detailed investigations will only be carried out once refinement is complete. The aim of this work is the rational design of mutations that will enhance the stability of proteins. Therefore mutations, designed by such comparative studies, are being created that should hopefully increase the stability of the mesophilic pig heart citrate synthase. The mutants will be able to be characterised with respect to the wild type by the experiments detailed in Chapter 4. Work is also in progress to rationally design mutants of the thermostable *Tp.acidophilum* citrate synthase that exhibit a higher thermostability than the wild type. This work will entail again detailed structural comparisons but with citrate synthases from hyperthermophilic Archaea (eg *Pyrococcus furiosus* which grows optimally at $100°C$).

The elucidation of more crystal structures for other Archaeal proteins should also allow identification of Archaeal protein features. Although at this early stage the separation of features contributing to thermal stability from possible archaeal features is difficult, features that are not observed in other thermostable proteins are possible candidates for such archaeal features.

Although crystal structures of thermostable proteins can yield vast amounts of information concerning the determinants of thermostability, two important facts must be remembered. Firstly, that a crystal structure is only a snap-shot of a dynamic protein and secondly, that the snap-shot has been taken at a sub-optimal temperature. To try and overcome these problems, two further complementary lines of research will be undertaken. Firstly, attempts will be made to grow crystals at higher temperatures (eg $55°C$) and to collect data at that temperature. Therefore the structure solved at room temperature will be able to be compared to that solved at a temperature where optimal activity is seen. Secondly, molecular dynamic simulations will be undertaken on both the pig heart and *Tp.acidophilum* citrate synthases at their respective working temperatures to investigate the role of flexibility in thermostability.

References

Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D. and Brown, F. (1989) Nature **337**, pp. 709-716

Alter, G. M., Casazza, J. P., Zhi, W., Nemeth, P., Srere, P. A. and Evans, C. T. (1990) Biochemistry **29**, pp. 7557-7563

Anderson, D. E., Becktel, W. J. and Dahlquist, F. W. (1990) Biochemistry **29**, pp. 2403-2408

Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frank, G. and Tratschin, J. D. (1979) Biochemistry **18**, pp. 5698-5703

Bacon, D. J. and Anderson, W. F. (1988) J. of Mol. Graphics **6**, pp. 219-220

Bhayana, V. and Duckworth, H. W. (1984) Biochemistry **23**, pp. 2900-2905

Bloxham, D. P., Parmlee, D. C., Kumar, S., Walsh, K. A. and Titani, K. (1982) Biochemistry **21**, pp. 2028-2036

Blundell, T. L. and Johnson, L. N. (1976) 'Protein Crystallography'. Academic Press, London.

Bradford, M. M. (1976) Anal. Biochem. **72**, pp. 248-252

Brady, L. and Jian-sheng, J. 'Molecular Replacement' – Proceedings of the CCP4Study Weekend, 31 January-1 February 1992 (Eds. E.J. Dodson, S. Gover, and W. Wolf), pp. 170-177. SERC, Daresbury.

Bright, J. R., Byrom, D., Danson, M. J., Hough, D. W. and Towner, P. (1993) Eur. J. Biochem. **180**. pp. 549-554

Brünger, A. T. (1990) Acta. Cryst. **A46**, pp. 46-57

Brünger, A. T., Kuriyan, J. and Karplus, M. (1987) Science **235**, pp. 458-460

Brünger, A. T. (1988) J. Mol. Biol. **203**, pp. 803-816

Burley, S. K. and Petsko, G. A. (1985) Science **229**, pp. 23-28

Carter, C. W., Baldwin, E. T. and Frick, L. (1988) J. Crystal Growth **90**, pp. 60-73

Castellano, E. E., Oliva, G. and Navaza, J. (1992) J. Appl. Cryst. **25**, pp. 281-284

CCP4 (1979). The SERC (U.K.) Collaborative computing project No. 4: A suite of programs for protein crystallography (Daresbury Lab., Warrington, U.K.)

Chang, C. T., Wu, C. -S. C. and Yang, J. T. (1978) Anal. Biochem. **20**, pp. 13-31

Christiansen, C., Freundt, E. A. and Black, F. T. (1975) Int. J. Syst. Bacteriol. **25**, pp. 99-101

Crowther, R. A. (1972) in 'The Molecular Replacement Method' (ed. M. G. Rossmann), pp. 174-178. New York: Gordon and Breach.

Crowther, R. A. and Blow, D. M. (1967) Acta. Cryst. **23**, pp. 544-548

Danson, M. J., Black, S. C., Woodland, D. L. and Wood, P. A. (1985) FEBS **179**, pp. 120-124

Dao-Pin, S., Sauer, U., Nicholson, H. and Matthews, B. W. (1991) Biochemistry **30**, pp. 7142-7153

Darland, G., Brock, T. D., Samonoff, W. and Conti, S. F. (1970) Science **170**, pp. 1416-1418

David, M., Lubinsky-Mink, S., Ben-Zvi, A., Suissa, M. and Ulitzur, S. (1991) Biochem. J. **278**, pp. 225-234

Davies, G. J., Gamblin, S. J., Littlechild, J. A. and Watson, H. C. (1993) Proteins: Structure, Function and Genetics **15**, pp. 283-289

Day, M. W., Hsu, B. T., Joshuator, L., Park, J. B., Zhou, Z. H., Adams, M . W. W. and Rees, D. C. (1992) Protein Science **1**, pp. 1494-1507

DeLange, R. J., Williams, L. C. and Searcy, D. G. (1981) J. Biol. Chem. **256**, pp. 905-911

Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acid Res. **12**, pp. 387-395

Donald, L. J. and Duckworth, H. W. (1987) Biochem. Cell Biol. **65**, pp. 930-938

Donald, L. J., Molgat, G. F. and Duckworth, H. W. (1989) J. Bacteriol. **171**, pp.5542-5550

Eijsink, V. G. H., Vriend, G., Vandervinne, B., Hazes, B., Vandenberg, B. and Venema, G. (1992) Proteins **14**, pp. 224-236

El-Kettani, M. A. E. C., Zakrzewska, K., Durup, J. and Lavery, R. (1993) Proteins **16**, pp. 393-407

Else, A. J., Danson, M. J. and Weitzman, P. D. J. (1988) Biochem. J. **254**, pp. 437-442

Eriksonn, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P. and Matthews, B. W. (1992) Science **255**, pp. 178-183

Evans, C. T., Owens, D. O., Sumegi, B., Kispal, G. and Srere, P. A. (1988) Biochemistry **27**, pp. 4680-4686

Fersht, A. R., Shi, J. P., Knill-Jones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M. and Brick, P. (1985) Nature **314**, pp. 235-238

Fitzgerald, P. M .D. (1988) J. Appl. Cryst. **21**, pp. 273-278

Fukaya, M., Takemura, H., Okumura, H., Kawamura, S., Horinouchi, S. and Beppu, T. (1990) J. Bacteriol. **172**, pp. 2096-2104

Gambacorta, A., Trincone, A. Nicolaus, B., Lama, L. and DeRosa, M. (1993) System. Appl. Microbiol. **16**, pp. 18-27

Gokhale, R. S., Agarwalla, S., Francis, V. S., Santi, D. V. and Balaram, P. (1994) J. Mol. Biol. **235**, pp. 89-94

Harper, E. T. and Rose, G. D. (1993) Biochemistry **32**, pp. 7605-7609

Heinzen, R. A., Frazier, M. E. and Mallavia, L. P. (1991) Gene **109** pp. 63-39

Horovitz, A., Matthews, J. M. and Fersht, A. R. (1992) J. Mol. Biol. **227**, pp. 560-568

Hough, D. W. and Danson, M. J. (1989) Letts. Appl. Microbiol. **9**, pp. 33-39

Huber, R. 'Molecular Replacement' – Proceedings of the Daresbury Study Weekend, 15-16 February 1985 (ed. P. A. Machin), pp. 58-61. SERC, Daresbury.

Huber, R. and Schneider, M. (1985) J. Appl. Cryst. **18**, pp. 165-169

Ikai, A. (1980) J. Biochem. **88**, pp. 1895-1898

Imada, K., Sato, M., Tanaka, N., Katsube, Y. Matsuura, Y. and Oshima, T. (1991) J. Mol. Biol. **222**, pp. 725-738

Ishikawa, K., Okumura, M., Katayanagi, K., Kimura, S., Kanaya, S., Nakamura, H. and Morikawa, K. (1993) J. Mol. Biol. **230**, pp. 529-542

Jack, A. and Levitt, M. (1978) Acta. Cryst. **A34**, pp. 931-935

Jaenicke, R. (1991) Eur. J. Biochem. **202**, pp. 715-728

James, K. D., Bonete, M. J., Byrom, D., Danson, M. J. and Hough, D. W. (1991) Biochem. Soc. Trans. **20**, 12S

James, M. N. G. and Sielecki, A. R. (1986) Nature **319**, pp. 33-38

Jancarik, J. and Kim, S. -H. (1991) J. Appl. Cryst. **24**, pp. 409-411

Jones, T. A. (1985) Methods in Enzymology **115** (Eds. H.W. Wycoff, C.H. Hirs, S.N. Timasheff), pp. 157-171

Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) Acta. Cryst. **A47**, pp. 110-119

Jones, W. T., Nayle Jr., D. P. and Whitman, W. B. (1987) Microbiol. Rev. **51**, pp. 136-177

Kabsch, W. (1988) J. Appl. Cryst. **21**, pp. 916-924

Kabsch, W. and Sander, C. (1983) Biopolymers **22**, pp. 2577-2637

Kandler, O. (1993) System. Appl. Microbiol. **16**, pp. 1-9

Karpusas, M., Holland, D. and Remington, S. (1991) Biochemistry **30**, pp. 6024-6031

Karpusas, M., Branchaud, B. and Remington, S. (1990) Biochemistry **29**, pp. 2213-2219

Kellis, J. T. Jr., Nyberg, K., Sali, D. and Fersht, A. R. (1988) Nature **333**, pp. 784-786

Kelly, C. A., Nishiyama, M., Ohnishi, Y., Beppu, T. and Birktoft, J. J. (1993) Biochemistry **32**, pp. 3913-3922

Kimura, S., Kanaya, S. and Makamura, H. (1992b) J. Biol. Chem. **267**, pp. 22014-22017

Kimura, S., Nakamura, H., Hashimoto, T., Oobatake, M. and Kanaya, S. (1992a) J. Biol. Chem. **267**, pp. 21535-21542

Kotik, M. and Zuber, H. (1993) Eur. J. Biochem. **211**, pp. 267-280

Kraulis, P. (1991) J. Appl. Cryst. **24**, 946-950

Kushner, D. J. (1985) in Woese, C. R. and Wolfe, R. S. (eds.): 'The Bacteria' **8**, pp. 171-214. Acad. Press Inc. (London) Ltd.

Lake, J. A. (1991) TIBS **16**, pp. 46-50

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L. and Pace, N. R. (1985) Proc. Natl. Acad. Sci. **82**, pp. 6955-6959

Laskowski, R. A., MacArthur, M. W., Moss, D. W. and Thornton, D. W. (1993) J. Appl. Cryst. **26**, pp. 283-291

Lattman, E. E. and Love, W. E. (1970) Acta. Cryst. **B26**, pp. 1854-1857

Liao, D-I., Karpusas, M. and Remington, S. J. (1991) Biochemistry **30**, pp. 6031-6036

Lill, U., Lefrank, S., Henschen, A. and Eggerer, H. (1992) **208**, pp. 459-466

Luo, M., Vriend, G., Kamer, G., Minor, I., Arnold, E., Rossmann, M. G., Boege, U., Scraba, D. G., Duke, G. M. and Palmenberg, A. C. (1987) Science **235**, pp. 182-191

Luthy, R., Bowie, J. U. and Eisenberg, D. (1992) Nature **356**, pp. 83-85

McPherson, A. (1990) Eur. J. Biochem. **189**, pp. 1-23

Man, W-J., Li, Y., O'Connor, D. and Wilton, D. C. (1991) **280**, pp. 521-526

Matsumura, M., Becktel, W. J., Levitt, M. and Matthews, B. W. (1989) Proc. Natl. Acad. Sci. **86**, pp. 6562-6566

Matthews, B. W. (1968) J. Mol. Biol. **33**, pp. 491-497

Matthews, B. W., Nicholson, H. and Becktel, W. J. (1987) Proc. Natl. Acad. Sci. **84**, pp. 6663-6667

McKenzie, H. A. and White, F. H. Jr. (1991) Adv. Prot. Chem. **41**, pp. 173-315

McEvily, A. J. and Harrison, J. H. (1986) J. Biol. Chem. **261**, pp. 2593-2598

Meiering, E. M., Serrano, L. and Fersht, A. R. (1992) J. Mol. Biol. **225**,

Menendez-Arias, L. and Argos, P. (1989) J. Mol. Biol. **206**, pp. 397-406

Merkler, D. J., Farrington, G. K. and Wedler, F. C. (1981) Int. J. Pept. Protein Res. **18**, pp. 430-442

Muir, J. M., Hough, D. W. and Danson, M. J. (1993) System. Appl. Microbiol. **16**, pp. 28-33

Nagai, K., Evans, P. R., Li, J. and Oubridge C. (1991) 'Isomorphous Replacement and Anomalous Scattering' – Proceedings of the CCP4Study Weekend, 25-26 January 1991 (Eds. W. Wolf, P. R. Evans and A. G. W. Leslie), pp. 141-149. SERC, Daresbury.

Navaza, J. (1987) Acta. Cryst. A**43**, pp. 645-653

Navaza, J. (1992) 'Molecular Replacement' – Proceedings of the CCP4 Study Weekend, 31 January-1 February 1992 (Eds. E.J. Dodson, S. Gover, and W. Wolf), pp. 87-90. SERC, Daresbury.

Navaza, J. (1993) Acta. Cryst. D**49**, pp. 588-591

Ner, S. S., Bhayana, V., Bell, A. W., Giles, I. G., Duckworth, H. W. and Bloxham, D. P. (1983) Biochemistry **22**, pp. 5243-5249

Numata, O., Takemasa, T., Takagi, I., Hirono, M., Hirano, H., Chiba, J. and Watanabe, Y. (1991) Biochem. Biophys. Res. Comm. **174**, pp.1028-1034

O'Neil, K. T. and DeGrado, W. F. (1990) Science **250**, pp. 646-651

Olsen, G. J. and Woese, C. R. (1989) Can. J. Microbiol. **35**, pp. 119-123

Olsen, G. J. and Woese, C. R. (1993) FASEB **7**, pp113-123

Patterson, A. L. (1934) Phys. Rev. **46**, pp. 372-376

Pechmann, H., Tesch, A. and Klink, F. (1991) FEMS Microbiol. Letts. **79.** pp. 51-56

Perutz, M. F. and Raidt, H. (1975) Nature **255**, pp. 256-259

Provencher, S. W. and Glöckner, J. (1981) Biochemistry **20**, pp. 33-37

Ree, H. K., Larsen, N., Gutell, R. and Zimmerman, R. A. (1993) System. Appl. Microbiol. **16**, pp. 333-341

Rehaber, V. and Jaenicke, R. (1992) J. Biol. Chem. **158**, pp. 10999-11006

Remington, S., Wiegand, G. and Huber, R. (1982) J. Mol. Biol. **158**, pp. 111-152

Remington, S. J. (1992) Curr. Topics Cell. Regul.**33**, pp. 209-229

Rennell, D., Bouvier, S. E., Hardy, L. W. and Poteete, A. R. (1991) J, Mol. Biol. **222**, pp. 67-87

Rosenkrantz, M., Alam, T., Kim K.-S., Clark, B. J., Srere, P. A. and Guarente, L. P. (1986) Mol. Cell. Biol. **6**, pp. 4509-4515

Rossmann, M. G. and Blow, D. M. (1962) Acta. Cryst. **15**, pp. 24-31

Schendel, F. J., August, P. R., Anderson, C. R., Hanson, R. S. and Flickinger, M. C. (1992) Appl. Environ. Microbiol. **58**, pp. 335-345

Schreuder, H. A., de Boer, B., Sixma, T. K., Tepliakov, A. V., Aguirre, A. and Hol, W. G. J. (1992) 'Molecular Replacement' – Proceedings of the CCP4Study Weekend, 31 January-1 February 1992 (Eds. E.J. Dodson, S. Gover, and W. Wolf), pp. 106-112. SERC, Daresbury.

Searcy, D. G. (1975) Biochim. Biophys. Acta. **395**, pp. 535-547

Searcy, D. G. and Doyle, E. K. (1975) Int. J. Syst. Bacteriol. **25**, pp. 286-289

Searcy, D. G. and Whatley, F. R. (1984) System. Appl. Microbiol. **5**, pp. 30-40

Searcy, D. G. (1986) Sys. Appl. Microbiol. **7**, pp. 198-201

Segerer, A., Langworthy, T.A. and Stetter, K.O. (1988) System. Appl. Microbiol. **10**, pp. 161-171

Serrano, L., Sancho, J., Hirshberg, M. and Fersht, A. R. (1992) J. Mol. Biol. **227**, pp. 544-559

Shirley, B. A., Stanssens, P., Hahn, U. and Pace, C. N. (1992) Biochemistry **31**, pp. 725-732

Siegel, J. R., Steinmetz, W. E. and Long, G. L. (1980) Anal. Biochem. **104**, pp. 160-167

Smith, L. D., Stevenson, K. J., Hough, D. W. and Danson, M. J. (1987) FEBS **225**, pp. 277-281

Srere, P. A., Brazil, H. and Gonen, L. (1963) Acta. Chem. Scand. **17**, pp. S129-S134

Stadtman, E. R. (1957) Meth. Enzymol. **13**, pp.931-932

Stein, D. B. and Searcy, D. G. (1978) Science **202**, pp. 219-221

Stetter, K. O. and Zillig, W. (1985) in Woese, C. R. and Wolfe, R. S. (eds.): 'The Bacteria' **8**, pp. 85-170. Acad. Press Inc. (London) Ltd.

Stuart, D. I. (1979) Ph.D. thesis. (Univ. of Bristol, U.K.).

Suissa, M. Suda, K. and Schatz, G. (1984) EMBO **3**, 1773-1781

Sutherland, K. J., Henneke, C. M., Towner, P., Hough, D. W. and Danson, M. J. (1990) FEBS **194**, pp. 839-844

Sutherland, K. J., Danson, M. J., Hough, D. W. and Towner, P. (1991) FEBS **282**, pp. 132-134

Unger, E. A., Hand. J. M., Cashmore, A. R. and Vasconcelos, A. C. (1989) Plant Mol. Biol. **13**, pp. 411-418

Wakabayashi, S., Fujimoto, N., Wada, K., Matsubara, H., Kerscher, L. and Oesterhelt, D. (1983) FEBS **162**, pp. 21-24

Walker, J. E., Wonacott, A. J. and Harris, J. I. (1980) Eur. J. Biochem. **108**, pp. 581-586

Wang, X and Janin, J. (1993) Acta. Cryst. **d49**, pp. 505-512

Weitzman, P. D. J. and Jones, D. (1968) Nature **219**, pp. 270-272

Weitzman, P. D. J. and Danson, M. J. (1976) Curr. Top. Cell. Regul. **10**, pp. 161-204

Weitzman, P. D. J. (1981) Adv. Microbiol. Physiol. **22**, pp. 185-244

Weitzman, P. D. J. and Ridley, J. (1983) Biochem. Biophys. Res. Comm. **112**, pp. 1021-1026

Wiegand, G., Remington, S., Deisenhofer, J. and Huber, R. (1984) J. Mol. Biol. **174**, pp. 205-219

Wiegand, G. and Remington, S. (1986) Annu. Rev. Biophys. Biophys. Chem. **15**, pp. 97-117

Wigley, D. B. 'Molecular Replacement' – Proceedings of the CCP4Study Weekend, 31 January-1 February 1992 (Eds. E.J. Dodson, S. Gover, and W. Wolf), pp. 145-153. SERC, Daresbury.

Winker, S. and Woese, C. R. (1991) System. Appl. Microbiol. **14**, pp. 305-310

Woese, C. R. and Fox, G. E. (1977) Proc. Natl. Acad. Sci. **74**, pp 5088-5090

Woese, C. R., Magrum, L. J. and Fox, G. E. (1978) J. Bacteriol. **169**, pp. 3564-3572

Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) Proc. Natl. Acad. Sci. **87**, pp. 4576-4579

Wood, D. O., Atkinson, W. H., Sikorski, R. S. and Winkler, H. H. (1983) J. Bacteriol. **155**, pp. 412-416

Wood, D. O., Williamson, L. R., Winkler, H. H. and Krause, D. C. (1987)

J. Bacteriol. **169**, pp. 3564-3572

Yang, D., Kaine, B. P. and Woese, C. R. (1985) System. Appl. Microbiol. **6**, pp. 251-256

Zhi, W., Srere, P. A. and Evans, C. T. (1991) Biochemistry **30**, pp. 9281-9286

Zillig, W., Klenk, H-P., Palm, P., Leffers, H., Puhler, G., Gropp, F. and Garrett, R. A. (1989) Endocytobiosis and Cell Res. **6**, pp. 1-25